TECHNISCHE UNIVERSITÄT DRESDEN

Fakultät Verkehrswissenschaften „Friedrich List“, Professur für Ökonometrie und Statistik, insb. im Verkehrswesen

Bachelorarbeit

Analyse des Pkw-Besitzes in Haushalten der 25 großen SrV-Vergleichsstädte

Stefan Martin Lins
Matrikelnummer:

Betreut durch

Dipl.-Verk.wirtsch. Stefanie Lösch

Verantwortlicher Hochschullehrer

Prof. Dr. Ostap Okhrin

Dresden, 06. September 2018

Zusammenfassung

In Zeiten des Klimawandels, erhöhten Feinstaubwerten, geänderten sozialen Wertevorstellungen und der Verfügbarkeit von Carsharing rückt der Pkw-Besitz in Haushalten immer wieder in den Fokus der Berichterstattung. Das Ziel dieser Arbeit ist es, Charakteristika zu finden, die den Pkw-Besitz beschreiben, und deren Wirkungen zu beurteilen. Der Besitz eines Pkws wird in der Literatur auf verschiedene Weise im Hinblick auf die Bedeutung als intervenierende Variable oder als Statussymbol untersucht. Als Grundlage dienen die Daten aus der Umfrage ‚SrV - Mobilität in Städten', wobei die Ergebnisse der 25 großen SrV-Vergleichsstädte verwendet werden. Diese Daten besitzen eine sogenannte Multilevelstruktur, das heißt, dass die Daten auf Wegeebene, Personenebene und Haushaltsebene separat vorliegen, wodurch eine Aggregation auf das Haushaltsniveau erforderlich wird. Der sich daraus ergebende Datensatz mit sozioökonomischen und alternativenspezifischen Variablen wird mithilfe deskriptiver Methoden sowie mit Zusammenhangsmaßen auf die Eignung als Variablen für die Anwendung des binären Logit-Modells untersucht, um aussagekräftige Ergebnisse generieren zu können. Mithilfe dieses Modells werden kardinale, kategoriale sowie nominale Variablen betrachtet und bewertet. Daraus lässt sich beispielsweise ableiten, dass der Pkw-Besitz in Haushalten mit zunehmender Personenzahl wahrscheinlicher ist, als bei Singlehaushalten. Auch das Einkommen und der fehlende Zugang zu Alternativen hat einen positiven Einfluss auf den Pkw-Besitz. Das Modell kann neben der Bestimmung der Eigenschaften dazu beitragen, den Pkw-Besitz in Haushalten zu prognostizieren. Interessant dabei ist, dass nicht alle Variablen die erwartete Wirkung entfalten. Die gefundenen Ergebnisse des Modells werden mit Erkenntnissen aus der Literatur verglichen, woraus sich einige Parallelen und Ergänzungen ergeben.

Abstract

Climate change, increasing fine dust, changes in values and the accessibility of Carsharing are well discussed topics nowadays in combination with the vehicle ownerships in German households. This paper aims to characterize the vehicle ownership and to evaluate their effects. National and international literature discusses the vehicle ownership in different ways like car ownership as status symbol or the variable 'vehicle ownership' as a mediating variable. Basis of this analysis is a survey called 'SrV - Mobilität in Städten'. The used data contains information about households in the 25 'großen SrV-Vergleichsstädte'. This information is available on different levels, which means that the information is available in separate datafiles for levels of ways, persons and households. The basis level for this analysis should be the household level. To get this level it is necessary to aggregate the information. As a result, we get several socioeconomic and alternative specific variables which must be investigated with descriptive and correlation methods in order to prove their suitability for the binary logit model. This model allows it to evaluate metric, nominal and categoric variables with the aim to find characteristics about vehicle ownership. Some results are for example that the vehicle ownership is more probable in households with more persons than in single-person households. Furthermore, the income and missing accessibility of alternatives have a positive effect on vehicle ownership. In addition, this model offers the possibility to predict the vehicle ownership in households. An interesting result is, that some variables have another effect than assumed. These results were compared with the findings of other papers. As a result, one can find some parallel and additional structures.

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

Abkürzungsverzeichnis

AIC	Akaike Informationskriterium
AUC	Area under Curve
BIC	Bayessches Informationskriterium
e.D.	Eigene Darstellung
L	Likelihood-Funktion
LL	Log-Likelihood-Funktion
LLR	Likelihood-Ratio-Statistik
LR-Test	Likelihood-Ratio-Test
OR	Odds-Ratio
ÖV	Öffentlicher Verkehr
SUV	Sports Utility Vehicle
PCC	Proportional Chance Criterion
Pkw	Personenkraftwagen
ROC	Receiver Operation Characteristic
RR	Relatives Risiko
SrV	System repräsentativer Verkehrsverhaltensbefragungen
VBA	Visual Basic for Applications

Symbolverzeichnis

A_{ol}	beobachtbare absolute Häufigkeit in Zelle o und Spalte l
$\tilde{A}_{ol}$	erwartete absolute Häufigkeit in Zelle o und Spalte l
a_i	Ausprägung des Merkmals i
b_j	Koeffizient des geschätzten Modells für die Variable j
C	Gesamtzahl der Beobachtungen
c	Laufindex des Beobachtungsfalls (c = 1, 2, ..., C)
D	Spaltenanzahl
d	Spaltenindex
E	Erwartungswert
e	Eulersche Zahl
$f(a_i)$	relative Häufigkeit der Merkmalsausprägung a_i
G	Anzahl der Kategorien der abhängigen Variablen
g	Laufindex für die Kategorien der Zufallsvariablen (g= 1,2 ..., G)
H_0	Nullhypothese
$h(a_i)$	absolute Häufigkeit der Merkmalsausprägung a_i
i	Laufindex für die Merkmalsausprägungen (i = 1,2, ..., I)
J	Anzahl der unabhängigen Variablen
j	Laufindex der unabhängigen Variablen (j = 1,2,...,J)
K	Kontingenzkoeffizient
K_{max}	Maximalwert für Kontingenzkoeffizient
K_*	Korrigierter Kontingenzkoeffizient
L_0	Likelihood des Nullmodells
L_v	Likelihood des vollständigen Modells
LL_0	Log-Likelihood des Nullmodells
LL_v	Log-Likelihood des vollständigen Modells
M	Minimum aus Zeilen- und Spaltenzahl
n	Gesamtzahl der Beobachtungen
O	Zeilenanzahl
o	Zeilenindex
P	Wahrscheinlichkeit
p	Wahrscheinlichkeit einer einzelnen Beobachtung
p^*	Trennwert
R^2	Bestimmtheitsmaß
R^2_{CS}	Cox & Snell-R^2
R^2_{McF}	McFadden's R^2

R_N^2	Nagelkerke's R^2
r	Korrelationskoeffizient
re_c	Residuum für Beobachtung c
s_{bj}	Standardfehler von b_j
t	Anteil einer der zwei Gruppen an der gesamten Zahl der Beobachtungen
u	Störterm
ue_c	Standardisiertes Residuum für Beobachtung c
v, w	Parameter für die Schätzung der Regressionsfunktion
W	Wald-Statistik
x	Ausprägung der unabhängigen Variablen
$x_{0,25}$	1. Quartil
$x_{0,5}$	Median
$x_{0,75}$	3. Quartil
x_{c_1}	Ausprägung der Variable 1 bei Beobachtung c
x_{c_2}	Ausprägung der Variable 2 bei Beobachtung c
$\bar{x}$	Mittelwert
$\bar{x}_1$	Mittelwert der Variablen 1 bei Beobachtung c
$\bar{x}_2$	Mittelwert der Variablen 2 bei Beobachtung c
Y	Abhängige Variable (Zufallsvariable)
y_c	Realisation des Ereignisses in Beobachtung c
z	Latente nicht erhobene Variable/systematische Komponente (Ergebnis aus Nutzenfunktion)
α	Signifikanzniveau
β_0	alternativenspezifische Konstante
β_j	Parameter für die Modellschätzung
π	Wahrscheinlichkeit der logistischen Regression
χ^2	Chi-Quadrat-Teststatistik
$>>$	viel größer als

1 Einleitung

Deutschland gilt im Vergleich zu anderen Ländern als eine Nation der Autofahrer, da der Anteil von Autonutzern im Gegensatz zu alternativen Fortbewegungsmitteln wie Öffentlicher Verkehr (ÖV), Fahrrad oder Zufußgehen wesentlich höher ist. Der wirtschaftliche Wohlstand in Deutschland ist nach wie vor stark mit dem Automobil verbunden und deutsche Automobilhersteller leisten seit der Erfindung des Autos Pionierarbeit bei Innovationen im Automobilsektor (vgl. Dienel 2007, S. 23). In enger Verbindung damit steht das Mobilitätsverhalten und der Personenkraftwagen (Pkw)-Besitz in Deutschland. Mobilität als Nachfrage nach Ortsveränderung stellt ein Grundbedürfnis des Menschen dar und geht mit der Nutzung von verschiedenen Verkehrsmitteln einher. Wird die Anzahl der beförderten Personen in Deutschland betrachtet, so wurden im Jahr 2016 75 % aller Personen mit einem Pkw und nur 17 % mit öffentlichen Verkehrsmitteln transportiert (vgl. BMVI 2017, S. 215). Der restliche Anteil verteilt sich auf die anderen Verkehrsmittel. Daraus wird bereits der Stellenwert des Pkws deutlich. Auf Haushaltsebene liegt der Anteil der verfügbaren Pkw bei 77,3 % mit leicht zunehmender Tendenz (vgl. ebd., S. 234). Die Entwicklung, dass in Deutschland immer mehr Personen über ein Auto verfügen, bestätigt auch der Motorisierungsgrad, der die Anzahl der Pkw pro 1.000 Personen angibt. Zwischen 2009 und 2017 stieg die Kennziffer der Motorisierung in Deutschland um fast 9 % auf 554 Pkw pro 1.000 Einwohner (vgl. ebd, S. 133, 327, Destatis 2018). Ordnet man diesen Wert in den europäischen Kontext ein, fällt auf, dass Deutschland bezogen auf das Jahr 2015 mit 548 Pkw pro 1.000 Einwohner deutlich über dem Durchschnitt der 28 EU-Länder von 498 Pkw pro 1000 Einwohner liegt (vgl. BMVI 2017, S. 327). Nur Finnland und Italien, die flächenmäßig etwa in der gleichen Größenordnung wie Deutschland liegen, weisen einen nochmals deutlich höheren Motorisierungsgrad auf (vgl. ebd.). Dies könnte auf mangelnde Alternativen zum eigenen Pkw zurückzuführen sein.

Vor dem Hintergrund des Klimawandels, steigendem Verkehrsaufkommen und zunehmender Feinstaubbelastung vorwiegend in Städten, mit denen der Pkw in einem kausalen Zusammenhang steht, ist ein näheres Verständnis des Pkw-Besitzes von besonderer Bedeutung. Die Dimensionierung von Verkehrsanlagen, das Vorhalten eines angemessenen Angebots im öffentlichen Verkehr und das Angebot an Alternativen zum Pkw können dazu beitragen, dem Trend zunehmender Motorisierung entgegen zu wirken. Hierfür ist es wichtig zu verstehen, welche Kriterien für den Pkw-Besitz in Haushalten von Bedeutung sind. Diese Kenntnisse können beispielsweise für die Dimensionierung neuer Wohngebiete von Nutzen sein. So können in neuen Wohnbaugebieten, in denen Familien leben sollen, andere Anforderungen an die Pkw-Stellplätze gestellt werden, als in Wohngebieten mit Einpersonenhaushalten. Zusätzlich können, anhand der Charakteristika, Ziele für den Verkehr in einem Wohngebiet aufgestellt werden. So ist es denkbar, dass ein Planungsorgan das Ziel eines geringen Pkw-Besitzes definiert. Dafür ist es wichtig, die Wirkung verschiedener Maßnahmen zu kennen, wie beispielsweise den Effekt von

Carsharing oder die Wirkung des Zugangs zu öffentlichen Verkehrsmitteln. Aus diesem Grund beschäftigt sich diese Arbeit mit der Analyse des Pkw-Besitzes in den Haushalten der 25 großen deutschen ‚System repräsentativer Verkehrsverhaltensbefragungen (SrV)-Vergleichsstädte' mit dem Ziel verschiedene Charakteristika für den Pkw-Besitz herauszufinden.

Die Arbeit ist folgendermaßen aufgebaut: Kapitel 2 beschäftigt sich mit einer Darstellung der aktuellen Literatur, die den Pkw-Besitz beschreibt. Die methodischen Grundlagen dieser Arbeit sind in Kapitel 3 enthalten. In Kapitel 4 werden die dieser Analyse zugrunde liegenden Daten beschrieben und deren Aufbereitung dargestellt. Kapitel 5 enthält die deskriptive Analyse der Merkmalsvariablen aus dem Datensatz. Die Anwendung und Interpretation der Ergebnisse des binären Logit-Modells ist in Kapitel 6 beschrieben. Daran schließt sich das Fazit (Kapitel 7) sowie der Literaturvergleich mit einer Diskussion der Ergebnisse (Kapitel 8) an. Abschließend werden die Ergebnisse in Kapitel 9 kritisch gewürdigt und ein Ausblick gegeben.

2 Literaturübersicht

Bereits in der Vergangenheit beschäftigten sich zahlreiche Autoren mit dem Thema Pkw-Besitz in Haushalten und dessen Einfluss auf das Mobilitätsverhalten.

Van Acker & Witlox (2010) untersuchen den Einfluss der Variable des Pkw-Besitzes als intervenierende Variable auf die persönliche Pkw-Nutzung. Die intervenierende Variable beschreibt den Einfluss anderer exogener Variablen auf den Pkw-Besitz und welche Wirkung wiederum der Pkw-Besitz als exogene Variable, neben den anderen exogenen Variablen, auf die Pkw-Nutzung hat. Laut Van Acker & Witlox (2010) werden in der bisherigen Literatur hauptsächlich die Einflüsse der baulichen Umgebung und der sozioökonomischen Faktoren auf das Verkehrsverhalten oder die Verkehrsmittelwahl untersucht. Der Pkw-Besitz hingegen spielt kaum eine Rolle. Die Autoren vergleichen verschiedene Modelle mit und ohne intervenierende Variable. Die Analyse der Zusammenhänge zwischen den Variablen und die Bedeutung der intervenierenden Variable wird über die Regressionsanalyse des Strukturgleichungsmodells durchgeführt. Die endogene Variable der Pkw-Nutzung ist dabei als kategoriale Variable mit den Ausprägungen ‚kein Pkw', ‚ein Pkw' und ‚mehr als ein Pkw' definiert. Beim Vergleich der Güte aller drei Modelle liefert das beste Ergebnis jenes Modell, welches den Pkw-Besitz als intervenierende Variable enthält. Die Autoren finden unter Anwendung dieses Modells heraus, dass der Pkw-Besitz in dicht bebauten Regionen geringer ist. Zudem sprechen Faktoren wie kurze Wege zur nächsten Bahnstation, niedriges Einkommen, hohes Alter, das Fehlen eines Führerscheines und Einpersonenhaushalte gegen den Pkw-Besitz. Andere Variablen wie Geschlecht, Bildung oder Kinder wurden aufgrund mangelnder Signifikanz aus dem Modell eliminiert. Weiterhin untersucht die Studie die Pkw-Nutzung basierend auf dem Pkw-Besitz. Dabei wird festgestellt, dass die Umgebung nur einen indirekten Einfluss auf die Pkw-Nutzung hat. Haupteinflussfaktoren auf den Pkw-Besitz sind laut Van Acker & Witlox (2010) das monatliche Haushaltseinkommen und der Führerscheinbesitz. Insgesamt wird ermittelt, dass einige Variablen keinen direkten Einfluss auf die Pkw-Nutzung haben, sondern eher auf den Pkw-Besitz, wie die Autoren am Beispiel des Einkommens zeigen. Lediglich die Wegelänge lässt sich direkt durch das Einkommen beeinflussen. Laut den Autoren wären Missspezifikationen zu erwarten, wenn die intervenierende Wirkung des Pkw-Besitzes nicht betrachtet werden würde. Der Pkw-Besitz hat folglich einen großen Einfluss als intervenierende Variable bei Betrachtung der Pkw-Nutzung.

Einen anderen Ansatzpunkt verfolgen Bhat & Sen (2006). Ziel deren Studie ist die Analyse der Eigenschaften von Haushalten mit einem Pkw, die Rückschlüsse auf einen bestimmten Pkw-Typen erlauben. Mithilfe eines multiplen diskret-stetigen Extremwertmodells untersuchen die Autoren den signifikanten Einfluss von soziodemografischen Variablen wie Einkommen, Existenz von Kindern, Haushaltsgröße, körperliche Einschränkungen oder das Geschlecht. Zudem fließen die Bevölkerungsdichte und die Betriebskosten der einzelnen Pkw-Typen in das Modell mit ein. In der Arbeit von Bhat & Sen (2006) werden als kategoriale Zufallsvariablen die ame-

rikanischen Pkw-Typen Passenger Car, Sports Utility Vehicle (SUV), Pickup Truck, Minivan und Van unterschieden. Die Autoren modellieren zudem die Veränderung des Pkw-Besitzes und der Pkw-Nutzung bei einer Zunahme der Kraftstoffpreise für alle Fahrzeugtypen. Eine Preissteigerung führe dazu, dass der Besitz großer Fahrzeugtypen in einem einstelligen Prozentbereich zurückginge, während die Anzahl der Passenger Cars nahezu stabil bleiben würde. Die Autoren finden in ihrer Arbeit heraus, dass mit zunehmender Anzahl an Kindern die Präferenz gegenüber SUVs und Minivans im Vergleich zu den Alternativen steigt. Minivans sind vor allem bei Haushalten mit mehreren Personen beliebt. Zudem bevorzugen abgelegene Haushalte in ländlichen Regionen und Haushalte mit männlichen Mitgliedern eher Pickup Trucks.

Die Bedeutung des Pkws als Statussymbol wird von Collin-Lange & Benediktsson (2011) für isländische Jugendliche im Alter von 16 bis 21 Jahren untersucht. In Island besitzt das Auto einen besonderen Stellenwert, denn gut neun von zehn volljährigen Isländern besitzen sowohl einen Führerschein als auch ein Auto. Die Autoren versuchen die Gründe dieser Entwicklung herauszufinden, indem sie eine Umfrage unter isländischen Hochschülern durchführten. Der Studie zufolge besitzt der größte Teil (98,2 %) dieser Schülergruppe einen Führerschein oder beabsichtigt, diesen zeitnah zu erwerben. Die Pkw-Nutzung unter den Schülern mit Führerschein ist mit 97 % sehr hoch, wobei 62 % bereits ein eigenes Auto besitzen. Selbst über 40 % der führerscheinlosen Schüler sind bereits im Besitz eines eigenen Pkws. Zwei von drei Schülern nutzen als Fahrer oder Mitfahrer einen Pkw auf dem Weg zwischen Wohnort und Schule. Busse werden gemäß der Studie nur bis zum Erwerb des Führerscheines benutzt bzw. danach vorwiegend von Schülern, die einen Schulweg von über 20 Kilometern haben. Begründen lässt sich dieser Pkw-Besitz laut den Autoren vordergründig durch das unzureichende Busangebot in Island, der höheren Flexibilität und durch zielloses Umherfahren, dem Rúntur. In der Studie gaben 77 % der Befragten an, dieser Art von Freizeitbeschäftigung nachzugehen. Dabei geht es vordergründig um soziale Kontakte, Freiheit von Zuhause und um die klassische Profilierung gegenüber Gleichaltrigen. Den Autoren zufolge bedeutet der Besitz eines Pkws in Island vor allem Freiheit und ist ein Resultat daraus, dass die jungen Erwachsenen das soziale und kulturelle Verhalten in ihrem Lebensumfeld nachahmen. Die Arbeit von Collin-Lange & Benediktsson (2011) stellt dabei deutlich den hohen Stellenwert des eigenen Pkws in Island als Statussymbol heraus und fordert dazu auf, diesen Trend als Anlass für die Entwicklung von Alternativen zu nehmen.

3 Methodik

3.1 Deskriptive Analyse

3.1.1 Lage- und Streumaße

Relative Häufigkeiten

Für die deskriptive Beschreibung von metrischen oder klassierten Daten empfiehlt sich eine Darstellung der relativen Häufigkeiten. Hierzu werden die vorkommenden Merkmalsausprägungen $a_i, ..., a_I$ für $i = 1, 2, ..., I$ bzw. die Klassen berücksichtigt. Die relative Häufigkeit $f(a_i)$ stellt das Verhältnis der absoluten Häufigkeit $h(a_i)$ zur Gesamtzahl der Beobachtungen n innerhalb einer Gruppe dar und lässt sich durch

$$f(a_i) = \frac{1}{n} * h(a_i) \tag{3.1}$$

berechnen. Zur Veranschaulichung der relativen Häufigkeit empfiehlt sich ein Säulendiagramm. (vgl. Bamberg, Baur & Krapp 2012, S. 11)

Boxplot

Für die Vergleichbarkeit der Verteilungen der erhobenen Werte zweier Kategorien eignet sich die Anwendung eines Boxplots (vgl. Schlittgen 2012, S. 33–35). Im Boxplot sind die in der folgenden Tabelle 3.1 genannten Werte zur Beurteilung enthalten.

Bezeichnung Lageparameter	**Formelzeichen**
Median	$x_{0,5}$
1. Quartil	$x_{0,25}$
3. Quartil	$x_{0,75}$
Mittelwert	$\bar{x}$

Tabelle 3.1: Elemente des Boxplots (e.D. in Anlehnung an Schlittgen (2012, S. 32))

Zusätzlich zu den oben genannten Elementen beinhaltet ein Boxplot sogenannte Whisker. In diesem Darstellungsmerkmal sind Daten innerhalb des 1,5-fachen Quartilsabstand zusammengefasst, ausgehend von den beiden Quartilen. Der Quartilsabstand bezeichnet dabei die Differenz zwischen dem 1. und 3. Quartil. Alle Elemente eines Datensatzes, die außerhalb dieser Linie liegen, werden als Ausreißer bezeichnet. (vgl. Schlittgen 2012, S. 33–35)

3.1.2 Zusammenhangsmaße

Korrigierter Kontingenzkoeffizient nach Karl Pearson

Dieser Abschnitt stützt sich auf Bamberg, Baur & Krapp (2012, S. 36 f.). Der Kontingenzkoeffizi-

ent K nach Karl Pearson beschreibt das Zusammenhangsmaß für zwei nominalskalierte Daten. Dieser wird mithilfe der als Chi-Quadrat (χ^2) bezeichneten Größe und der Grundgesamtheit berechnet. χ^2 ergibt sich aus der Summe der quadrierten Differenz zwischen der beobachteten absoluten Häufigkeit A_{od} in Zeile o und Spalte d sowie der entsprechend erwarteten absoluten Häufigkeit $\tilde{A}_{od}$ dividiert durch $\tilde{A}_{od}$, wie folgende Formel beschreibt

$$\chi^2 = \sum_{o=1}^{O} \sum_{d=1}^{D} \frac{(A_{ol} - \tilde{A}_{od})^2}{\tilde{A}_{od}}. \tag{3.2}$$

Ausdrücken lässt sich der Kontingenzkoeffizient folgendermaßen

$$K = \sqrt{\frac{\chi^2}{n + \chi^2}}, \tag{3.3}$$

wobei ein Wert nahe der null keinen oder einen geringen Zusammenhang zwischen den Variablen widerspiegelt. Mit einem zunehmenden Wert von χ^2 wächst K asymptotisch dem Wert eins entgegen, der allerdings aufgrund der Tabellengröße nicht erreicht werden kann, was von der gesamten Zeilenzahl O und Spaltenzahl D abhängt. Der Maximalwert für den Kontingenzkoeffizient K_{max} lässt sich berechnen durch

$$K_{max} = \sqrt{\frac{M-1}{M}} \text{ wobei, } M = min\{o, d\}. \tag{3.4}$$

Um ein Normierungsintervall von [0; 1] zu erreichen, ist die Berechnung des korrigierten Kontingenzkoeffizienten K_* erforderlich:

$$K_* = \frac{K}{K_{max}}. \tag{3.5}$$

Je höher der Wert für K_* ist, desto höher ist der Zusammenhang zwischen zwei nominalen Variablen.

Korrelationstabelle

Nach Backhaus et al. (2016, S. 393 f.) können verschiedene unterschiedliche Variablen innerhalb eines Modells Zusammenhänge untereinander aufweisen und damit die Güte des Modells beeinflussen. Diese lineare Abhängigkeit zwischen unabhängigen Variablen wird als Multikollinearität bezeichnet (vgl. Albers et al. 2009, S. 221). Aus diesem Grund besteht die Erfordernis, diese Zusammenhänge zwischen den Ausgangsvariablen messbar zu machen, um diese bewerten zu können (vgl. Backhaus et al. 2016, 393 f.). Der Korrelationskoeffizient r zwischen zwei unabhängigen, metrischen Variablen lässt sich über die folgende Formel berechnen

$$r_{x_1,x_2} = \frac{\sum_{c=1}^{C} (x_{c_1} - \bar{x_1}) * (x_{c_2} - \bar{x}_2)}{\sqrt{\sum_{c=1}^{C} (x_{c_1} - \bar{x}_1)^2 * (x_{c_2} - \bar{x}_2)^2}}, \tag{3.6}$$

wobei

x_{c_1} = Ausprägung der Variable 1 bei Beobachtung c,
$\bar{x}_1$ = Mittelwert der Ausprägung von Variable 1 über alle Beobachtungen c,
x_{c_2} = Ausprägung der Variable 2 bei Beobachtung c,
$\bar{x}_2$ = Mittelwert der Ausprägung von Variable 2 über alle Beobachtungen c (vgl. ebd.).

Je höher der Wert für den Korrelationskoeffizienten, desto größer ist der Zusammenhang zwischen den beiden Variablen (vgl. ebd.).

3.2 Binäre logistische Regression

3.2.1 Allgemeines

Dieser Abschnitt bezieht sich - sofern nicht anders vermerkt - auf Backhaus et al. (2016, S. 284–334). Die logistische Regression lässt sich zur Gruppe der strukturen-prüfenden Verfahren zuordnen. Damit können Fragestellungen beantwortet werden, die Aussagen über zwei oder mehrere alternative Zustände treffen und damit auf das wahrscheinlichste Ergebnis schließen lassen. Diese besondere Form der Regressionsanalyse zeichnet sich dadurch aus, dass die abhängige Variable Y eine kategoriale Variable ist, wobei es sich bei den Ausprägungen (g = 1,...,G) um die verschiedenen Alternativen handelt. Vor dem Hintergrund des unsicheren Eintretens der Ereignisse stellt Y eine Zufallsvariable dar. Für die Zuordnung zu den Ausprägungen von Y werden Wahrscheinlichkeiten der einzelnen Zustände prognostiziert. Bei zwei abhängigen Gruppierungsvariablen wird von binärer logistischer Regression und bei mehr als zwei wird von multinomialer logistischer Regression gesprochen. Darüber hinaus wird zwischen einfacher und multipler logistischer Regression unterschieden. Erstere ist gegeben, wenn es lediglich eine erklärende Variable gibt, während im anderen Fall mindestens zwei erklärende Variablen vorliegen.

Im Gruppenfall G=2 wird die abhängige Variable binär, wobei die Wahrscheinlichkeiten P in der Form

$$P(Y = 0) = 1 - P(Y = 1) \tag{3.7}$$

dargestellt werden können, wenn Y die Werte null oder eins annimmt.

Grundsätzlich lässt sich die logistische Regression als

$$\pi(x) = f(x_1, ..., x_J) \tag{3.8}$$

darstellen und der Ausdruck $\pi(x) = P(Y = 1|x)$ steht für die bedingte Wahrscheinlichkeit, wenn Ereignis 1 für vorhandene Werte der unabhängigen Variablen $x_1, ..., x_J$ eintritt, wobei eine lineare Kombination der unabhängigen Variablen vorgenommen wird. Die systematische Komponente dieses Modells wird mithilfe der Linearkombination

$$z_c(x) = \beta_0 + \sum_{j=1}^{J} \beta_j * x_{jc} + u_c, \tag{3.9}$$

wobei

z = latente nicht erhobene Variable,
β_0 = alternativenspezifische Konstante,
β_j = Parameter für Modellschätzung,
x = Ausprägung der unabhängigen Variablen,
j = Laufindex der unabhängigen Variablen (j=1,2,...,J),
c = Index des Beobachtungsfalls (c=1,2,...,C),
u = Störterm

beschrieben und ist in diesem Fall gleichzusetzen mit der linearen Regressionsanalyse (vgl. Albers et al. 2009, S. 267 f.). Der Wert für z steht für den Gesamtnutzen, der sich in deterministischen, also beobachtbaren Nutzen und Zufallsnutzen aufteilt. Unter der Annahme, dass der Zufallsnutzen, also die nicht beobachtbaren Ergebnisse im Störterm, unabhängig identisch Gumbel-verteilt ist, lässt sich das Logit-Modell ableiten (vgl. Maier & Weiss 1990, S. 135 f.). Dieses Vorgehen wird im Folgenden mit der logistischen Funktion weiter beschrieben.

Anders als die lineare Regression basiert die logistische Regression auf der logistischen Funktion

$$p = \frac{e^z}{1+e^z} = \frac{1}{1+e^{-z}}. \tag{3.10}$$

Mithilfe dieser lassen sich Variablen mit reellen Werten von $[-\infty;\infty]$ in den Wahrscheinlichkeitsbereich $[0;1]$ transformieren. Unter Anwendung der Transformation der systematischen Komponente mit der logistischen Funktion lässt sich die logistische Regressionsfunktion

$$\pi(x) = \frac{1}{1+e^{-z(x)}} \tag{3.11}$$

mit einem S-förmigen Verlauf abbilden, wobei die systematische Komponente $z(x)$ ein Prädikator für die Wahrscheinlichkeit $\pi(x)$ ist. Bei zunehmenden $z(x)$ wächst ebenso $\pi(x)$, was wiederum bedeutet, dass je größer die systematische Komponente ist, desto kleiner ist der Wert für $P(Y = 0|x)$.

Im folgenden Abschnitt wird die Vorgehensweise zur Ermittlung des Modells, die Prüfung des Gesamtmodells sowie die Prüfung der Merkmalsvariablen erläutert.

3.2.2 Modellformulierung

Im Vorfeld der Modellermittlung sind verschiedene Schritte zur Modellerstellung vorzunehmen. In erster Linie sind die abhängigen Variablen und die erklärenden Variablen zu definieren. Hierzu müssen die alternativ möglichen Zustände im Hinblick auf das Ziel des Modells bzw. der Analyse ermittelt werden. Zudem müssen anhand von hypothetischen Annahmen die erklärenden Variablen festgelegt werden, die im Rahmen von theoretischen oder sachlogischen Überlegungen zu ermitteln und durch deskriptive Analysen zu beurteilen sind. Bei binären Modellen wird häufig Y = 1 für den Eintritt eines Ereignisses und Y = 0 für die Ablehnung des Ereignisses gesetzt. Im binären logistischen Regressionsmodell wird für die Zufallsvariable, die zwei Werte annehmen

kann, unterstellt, dass diese unabhängig verteilt ist mit dem Erwartungswert $E(Y_c|x_c) = \pi(x_c)$. In diesem Fall wird von Bernoulli-Variablen gesprochen, sodass die dem Modell zugrunde liegende Wahrscheinlichkeitsverteilung die Bernoulli-Verteilung ist.

Die multiple logistische Regression lässt sich unter Vernachlässigung des Störterms, aufbauend auf den zuvor erwähnten Formulierungen, in der Form

$$\pi(x_c) = \frac{1}{1 + e^{\beta_0 + \beta_1 x_{1c} + ... + \beta_J x_{Jc}}} \tag{3.12}$$

darstellen. $x_c = (x_{1c}, x_{2c}, ..., x_{Jc})$ spiegelt dabei die Werte der unabhängigen Variablen einer Beobachtung c wider. Die Eintrittswahrscheinlichkeit $\pi(x_c)$ erfährt eine Determination durch x_c und lässt sich wie folgt in allgemeiner Weise ausdrücken:

$$\pi(x_c) = P(Y_c = 1|x_c). \tag{3.13}$$

3.2.3 Schätzung der logistischen Regressionsfunktion

Die logistische Funktion weist die Eigenschaft der Nichtlinearität auf, sodass zur Schätzung der Regressionskoeffizienten die Maximum-Likelihood-Methode zur Anwendung kommt. Hiernach sind für die unabhängigen Parameter Schätzungen vorzunehmen mit dem Ziel, dass die realisierten Daten höchstmögliche Plausibilität (Likelihood) erreichen. Beim logistischen Regressionsmodell kann dies insofern interpretiert werden, dass für eine Beobachtung c die Wahrscheinlichkeit $p(x_c)$ den größtmöglichen Wert annehmen soll, wenn der Wert für die tatsächliche Realisation des Ereignisses der Beobachtung y_c gleich eins bzw. null ist.

Zusammengefasst lässt sich dies durch folgenden Ausdruck darstellen, der so groß wie möglich sein sollte:

$$p(x_c)^{y_c} * [1 - p(x_c)]^{1-y_c}. \tag{3.14}$$

Eine Annahme des Modells ist, dass Y_c über alle Beobachtungen hinweg eine voneinander unabhängige Verteilung aufweisen sollen, womit sich die gemeinsame Wahrscheinlichkeit als Produkt der Einzelwahrscheinlichkeiten ausdrücken lässt. Daraus folgt die zu maximierende Likelihood-Funktion (L):

$$L(v, w) = \prod_{c=1}^{C} p(x_c)^{y_c} * [1 - p(x_c)]^{1-y_c} \rightarrow Max!, \tag{3.15}$$

wobei $y_c = 1$ für bzw. $y_c = 0$ gegen den Eintritt eines Ereignisses steht.

Die Parameter v und w sollen so geschätzt werden, dass die Likelihood-Funktion ihr Maximum annimmt. Für die Schätzung wird eine vereinfachte Funktion verwendet, bei der die Wahrscheinlichkeiten logarithmiert werden und eine Umwandlung des Produkts in eine Summe stattfindet. Damit ergibt sich die sogenannte Log-Likelihood-Funktion (LL) mit dem Ausdruck:

$$LL(v, w) = \sum_{c=1}^{C} ln[p(x_c)] * y_c + ln[1 - p(x_c)] * (1 - y_c) \rightarrow Max!. \tag{3.16}$$

Mit der Eigenschaft einer streng monoton steigenden Funktion, ergibt sich durch Maximie-

rung beider vorangegangener Ausdrücke der gleiche Wert. Weiterhin weist die LL-Funktion die Eigenschaft auf, dass diese nur negative Werte annehmen kann, sodass eine Maximierung mit einem Wert null als Ziel vorgenommen wird. Dies würde bedeuten, dass das Modell in seiner Ausgangslage alle Werte der richtigen Kategorie zuordnet. Für die Maximierung bedient sich die LL-Funktion iterativer Algorithmen wie dem Quasi-Newton-Verfahren, wozu die konvexe Führung der LL-Funktion einen positiven Einfluss hat, da damit lokale Optima nicht berücksichtigt werden.

3.2.4 Prüfung des Gesamtmodells

Nachdem die einzelnen Koeffizienten mithilfe der Maximum-Likelihood-Methode geschätzt wurden, gilt es die Güte des Modells zu beurteilen. Hier sind verschiedene Verfahren möglich, die teilweise unterschiedliche Ergebnisse liefern. Dabei spielt die Anzahl der verwendeten Variablen für die Schätzung eine bedeutende Rolle. Zur Prüfung des geschätzten Modells liegt es nahe, den Ausgangsdatensatz in eine Lernstichprobe und eine Kontrollstichprobe aufzuteilen sowie an den Testdaten eine Klassifizierung vorzunehmen.

Log-Likelihood-Wert

Die angewandte Methode für die Schätzung der Regression ist die unter 3.2.3 dargestellte Maximum-Likelihood-Methode. Der zur Schätzung verwendete Maximalwert der LL kann ebenfalls als Basis zur Beurteilung der Güte von verschieden Modellen herangezogen werden. Zur Bewertung der Modelle wird der Ausdruck $-2LL$ verwendet, wobei die Multiplikation mit dem Faktor 2 darin begründet ist, dass eine Chi-Quadrat-verteilte Teststatistik anzustreben ist. Die Transformation in den positiven Bereich folgt daraus, dass der negative LL-Wert positiv wird. Je besser ein Modell ist, desto niedriger ist der Wert für $-2LL$, woraus folgt, dass ein Modell mit einem niedrigeren Wert vorzuziehen ist.

Informationskriterien

Mit zunehmender Anzahl an unabhängigen Variablen in einem Modell wird der Wert der LL ebenfalls kleiner. Dies ist grundsätzlich positiv zu beurteilen. Allerdings ist damit verbunden, dass sich das Modell immer mehr an die konkrete Stichprobe anpasst, was im Hinblick auf die Repräsentation der Grundgesamtheit nicht zielführend ist. Aus diesem Grund ergibt sich das wichtige Kriterium der Sparsamkeit (model parsimony) für die Modellschätzung. Bedeutende Kriterien zur Beurteilung der Modellgüte, insbesondere auch im Vergleich verschiedener Modelle, sind das Akaike Informationskriterium (AIC) und das Bayessches Informationskriterium (BIC). Diese Gütekriterien wirken sich insofern aus, dass sich eine zunehmende Anzahl an Parametern nachteilig auf den Kriteriumswert auswirkt. Die Kriterien lassen sich in folgender Weise berechnen:

$$AIC = -2 * LL + 2 * \text{Zahl der Parameter} \tag{3.17}$$

$$BIC = -2 * LL + ln(C) * \text{Zahl der Parameter} \tag{3.18}$$

$$\text{Zahl der Parameter (Freiheitsgrade)} = [(G-1)(J+1)] \tag{3.19}$$

mit

J = Anzahl der unabhängigen Variablen,
G = Anzahl der Kategorien der abhängigen Variablen.

Für die Beurteilung ist es von Bedeutung, dass das Modell mit dem niedrigsten Wert das beste Ergebnis liefert, wobei sich der Wert für das AIC und BIC bei der Beurteilung unterscheiden kann.

Klassifizierungstabelle

Nach der Schätzung des Modells können anhand der geschätzten Wahrscheinlichkeiten die einzelnen Datensätze zur Prognose der Zuordnung genutzt werden. Dies erfolgt unter Festlegung eines Trennwertes (p^*) für den gilt:

$$y_c = \begin{cases} 1, & \text{wenn } p_c > p^* \\ 0, & \text{wenn } p_c \leq p^* \end{cases} . \tag{3.20}$$

Die Klassifizierung als Gütemaß wird für zurückliegende Daten vorgenommen. Dies erfolgt mithilfe einer Klassifizierungstabelle (Confusion-Matrix). Diese ist hauptsächlich in vier Felder aufgeteilt, wobei zwischen Treffer und Nicht-Treffer für die einzelnen Gruppen unterschieden wird. Die Diagonale enthält dabei Informationen über die richtig zugeordneten Datensätze, woraus sich die Trefferquote errechnen lässt. Dieses Gütemaß enthält den Anteil der richtig klassifizierten Fälle an der Gesamtzahl der Fälle. Ein weiteres Gütemaß ist die Sensitivität, die den Anteil der richtigen Prognosen an der Gesamtzahl des Eintritts der tatsächlichen Ereignisse widerspiegelt. Das Gegenstück hierzu ist die Spezifität, die die richtig prognostizierten Nicht-Ereignisse mit deren tatsächlicher Gesamtzahl ins Verhältnis setzt.

Für eine sachgemäße Beurteilung der Trefferquote ist ein Vergleich mit einer zufälligen Zuordnung der Datenelemente notwendig. Diese liegt bei gleicher Gruppengröße bei 50 %. Bei bekannten und ungleich großen Stichprobenzahlen entspricht der Anteil der richtig klassifizierten Elemente, im Falle der zufälligen Zuordnung, dem Anteil der größten Gruppe am gesamten Umfang der Stichprobe. Unter dem Aspekt, dass die Zufallszuordnung stark von der Zahl der Gruppenelemente abhängt, wird mithilfe des Proportional Chance Criterion (PCC) der Anteil der Trefferquote bei Zuordnung aller Elemente zur größeren Gruppe berechnet, der durch die Trefferquote übertroffen werden sollte (vgl. Albers et al. 2009, S. 271). Der folgende Ausdruck versucht Verzerrungen zu vermeiden

$$PCC = t^2 + (1 - t)^2, \tag{3.21}$$

wobei

t = Anteil einer der zwei Gruppen an den gesamten Beobachtungen (vgl. ebd.).

Daraus ergibt sich, dass die Trefferquote aus der Klassifizierungstabelle einen höheren Wert annehmen soll, als sich nach dem Zufallsprinzip ergeben würde. Als Stichprobeneffekt wird der Fall bezeichnet, wenn die Trefferquote aus der Lernstichprobe errechnet wird. Bei Anwendung des Kontrolldatensatzes ist eine niedrigere Trefferquote als sehr wahrscheinlich anzunehmen.

Dieser Effekt wird jedoch mit zunehmender Anzahl an Beobachtungen geringer (vgl. Backhaus et al. 2016, S. 238 f.).

ROC-Kurve

Allgemeiner als die Klassifizierungstabelle, die jeweils nur für eine bestimmte Trefferquote gilt, ist die Receiver Operation Characteristic (ROC)-Kurve. Die ROC-Kurve fasst alle Klassifizierungstabellen aller möglichen Trennwerte zusammen. Ein Punkt auf der ROC-Kurve steht für einen bestimmten Trennwert, wobei die Ordinate die Sensitivität und die Abszisse die 1-Spezifität darstellt. Zudem enthält das Diagramm eine Winkelhalbierende, die bei einer rein zufälligen Zuordnung zu erwarten wäre. Als Maß für die Güte der Prognose- bzw. Klassifizierungsfähigkeit kann die Area under Curve (AUC) verwendet werden. Diese lässt sich wie folgt ausdrücken

$$AUC = \int_0^1 f(x)dx, \tag{3.22}$$

wobei $f(x)$ für die Funktion der ROC-Kurve steht. Ab einem Wert von 0,7 kann von einer akzeptablen Güte des Modells gesprochen werden.

LR-Test

Der Likelihood-Ratio-Test (LR-Test), der auch als Likelihood-Quotienten-Test bekannt ist, stellt den wichtigsten Test für die Güteprüfung eines Logit-Modells dar. Dieser vergleicht den LL-Wert des vollständigen Modells mit dem des Nullmodells. Im Nullmodell werden keine Variablen berücksichtigt, sondern nur die Konstante errechnet. Es hat demnach die schlechteste Anpassung an die Realität. Die Likelihood-Ratio-Statistik (LLR) lässt sich folgendermaßen ausdrücken:

$$LLR = -2 * ln\left(\frac{\text{Likelihood des Nullmodells}}{\text{Likelihood des vollständigen Modells}}\right) \tag{3.23}$$

$$= -2 * ln\left(\frac{L_0}{L_v}\right) = -2 * (LL_0 - LL_v) \tag{3.24}$$

mit

LL_0: Maximierte LL für das Nullmodell

LL_v: Maximierte LL für das vollständige Modell.

Daraus lässt sich die Nullhypothese H_0: $\beta_1 = \beta_2 = ... = \beta_J = 0$ bilden, unter der der Wert für LLR χ^2-verteilt ist mit J Freiheitsgraden. Mithilfe dieser lässt sich die Signifikanz unter einem gewissen Signifikanzniveau α des Modells feststellen.

Pseudo-R-Quadrat-Statistiken

Im Gegensatz zur linearen Regression, bei der die abhängige Variable metrisch ist, kann bei der logistischen Regression kein Bestimmtheitsmaß R^2 bestimmt werden, das den Anteil der Streuung der abhängigen Variable durch das Modell angibt. Als Ersatz hierfür werden in der logistischen Regressionsanalyse sogenannte Pseudo-R^2-Statistiken verwendet, die sich im Wertebereich von null bis eins bewegen und insofern beurteilt werden können, dass ein größerer Wert eine bessere Anpassung gewährleistet. Im Gegensatz zum klassischen Bestimmtheitsmaß wird

nicht der Anteil der erklärten Streuung an der Gesamtheit der Streuung erklärt, sondern das Wahrscheinlichkeitsverhältnis vom Nullmodell zum vollständigen Modell, analog zur LLR.

McFadden's R^2

Mithilfe des McFadden's R^2 wird der Quotient des Log-Likelihoods gebildet, wodurch sich dieser von der LLR unterscheidet. Das McFadden's R^2 lässt sich wie folgt ausdrücken:

$$R^2_{McF} = 1 - \left(\frac{LL_v}{LL_0}\right). \tag{3.25}$$

Liegt das Nullmodell nahe dem vollständigen Modell, entspricht der Quotient nahezu eins und das McFadden's R^2 wird sich null annähern. Ein Wert von eins ist bei realen Datensätzen kaum zu erreichen, da dieser einem Log-Likelihood von null entsprechen würde und eine perfekte Anpassung impliziert.

Cox & Snell-R^2

Ein weiteres Gütemaß stellt das Cox & Snell-R^2 dar, mit

$$R^2_{CS} = 1 - \left(\frac{L_0}{L_v}\right)^{\frac{2}{C}}, \tag{3.26}$$

wobei dessen Werte nur unter eins liegen können, unter der Annahme, dass L_0 immer Werte größer null annimmt. So werden auch bei perfekter Anpassung Werte kleiner als eins geliefert.

Nagelkerke's R^2

Der Gütetest von Nagelkerke wird auf Basis der Statistik von Cox & Snell gebildet und modifiziert diese so, dass auch der Maximalwert von eins angenommen werden kann. Die Formel für Nagelkerke's R^2 lautet:

$$R^2_N = \frac{R^2_{CS}}{1 - L_0^{2/C}}. \tag{3.27}$$

Eine Einordnung der Gütemaße für die Güteprüfung kann Tabelle 3.2 entnommen werden.

Gütemaß	**Wertebereiche**
Likelihood-Ratio-Test	hoher χ^2-Wert; Signifikanzniveau $< 5\%$
McFadden	$> 0{,}2$ akzeptabel $\mid > 0{,}4$ gut
Cox & Snell	$> 0{,}2$ akzeptabel $\mid > 0{,}4$ gut
Nagelkerke	$> 0{,}2$ akzeptabel $\mid > 0{,}4$ gut $\mid > 0{,}5$ sehr gut
Klassifikationsmatrix	Trefferquote $>$ PCC

Tabelle 3.2: Gütemaße und deren Wertebereiche (e.D. in Anlehnung an Albers et al. (2009, S. 272))

3.2.5 Prüfung der Merkmalsvariablen

Im Rahmen der Güteprüfung der Merkmalsvariablen wird überprüft, ob die unabhängigen Variablen einen signifikanten Einfluss auf die Zufallsvariable haben und damit deren Koeffizient

signifikant von null abweicht. Im Rahmen der logistischen Regression wird der Wald-Test und der Likelihood-Ratio-Test verwendet.

Wald-Test

Die Wald-Statistik, mit der über die Ablehnung oder Annahme einer Nullhypothese entschieden wird, lässt sich mit der Formel

$$W = \left(\frac{b_j}{s_{bj}} \right)^2 \tag{3.28}$$

darstellen, wobei

s_{bj} = Standardfehler von b_j.

Dabei ist die Wald-Statistik mit der Nullhypothese H_0: $\beta_i = 0$ mit einem Freiheitsgrad asymptotisch χ^2-verteilt.

Likelihood-Ratio-Test

Ebenso wie bei der im Abschnitt 3.2.4 dargestellten Anwendung des Likelihood-Ratio-Tests bei der Modellprüfung, kann dieser für die Prüfung der Merkmalsvariablen angewandt werden. Es wird dabei der Likelihood des vollständigen (LL_v) dem des reduzierten Modells gegenübergestellt, welches sich dadurch ergibt, dass der zu prüfende Koeffizient b_j auf null gesetzt wird und danach die Maximierung der restlichen Parameter erfolgt. Der sich ergebende Maximalwert wird als LL_{0j} bezeichnet. Der Wert für die Beurteilung lässt sich durch

$$LLR_j = -2 * (LL_{0j} - LL_v) \tag{3.29}$$

berechnen und wird als Likelihood-Statistik bezeichnet. Auch hier ist der Wert für LLR_j unter der Nullhypothese H_0: $b_j = 0$ asymptotisch χ^2-verteilt mit einem Freiheitsgrad.

Werden beide Prüfungsmodelle verglichen, so weist der Wald-Test systematisch, aufgrund des zu großen Standardfehlers bei großen absoluten Schätzwerten, einen größeren p-Wert auf, weshalb der aufwendigere LR-Test vorzuziehen ist. Bei großen Stichproben gleichen sich die Ergebnisse der Modelle einander an.

3.2.6 Residuen-Analyse

Die logistische Regression hat die Eigenschaft relativ unempfindlich auf Ausreißer zu reagieren. Eine Kontrolle auf Ausreißer ist trotzdem zielführend, um möglicherweise fehlerhafte Werte zu identifizieren oder deren Plausibilität darstellen zu können. Über die Residuuen re_c, die der Differenz zwischen beobachteten Wert und der geschätzten Wahrscheinlichkeit gemäß der Formel

$$re_c = y_c - p_c, \tag{3.30}$$

entsprechen, lassen sich Ausreißer ermitteln. Hier gilt, dass die Summe der Residuen gleich null ist.

Für aussagekräftigere Analysen sind die standardisierten Residuen ue_c heranzuziehen, die den Quotienten aus den Residuen und der Standardabweichung der Bernoulli-Verteilung entsprechen:

$$ue_c = \frac{y_c - p_c}{\sqrt{p_c(1-p_c)}}. \tag{3.31}$$

Bei einem großen Stichprobenumfang C unterliegen die standardisierten Residuen, auch Pearson-Residuen bezeichnet, annähernd der Normalverteilung mit einem Mittelwert von null und einer Standardabweichung von eins.

3.2.7 Interpretation der Regressionskoeffizienten

Unter dem Aspekt des nichtlinearen Verlaufs logistischer Regressionen bedarf die Interpretation der Koeffizienten einer besonderen Betrachtung. Dies beruht darauf, dass die Wirkungen der Koeffizienten keine Konstanz aufweisen. Aus diesem Grund lassen sich nur Aussagen über die Änderung der abhängigen Variablen treffen, basierend auf einer Variation der unabhängigen Variablen. Eine Aussage über die Stärke der Veränderung lässt sich wiederum nicht treffen. Zur einfacheren Interpretation der Koeffizienten im Modell der logistischen Regression werden häufig Odds und Logits verwendet, welche in den folgenden Absätzen beschrieben werden.

Odds

Unter Odds werden Chancen verstanden, die sich aus dem Quotienten der Wahrscheinlichkeit zu ihrer Gegenwahrscheinlichkeit ergeben und sich in der Form

$$odds = \frac{p}{1-p} = e^{z(x)} \tag{3.32}$$

darstellen lassen. Hierbei wird die Chance für eine einzelne Beobachtung bewertet. Die Odds besitzen die Eigenschaft, dass sie immer einen positiven Wert annehmen und keine obere Grenze aufweisen.

Für die Interpretation des logistischen Modells von hervorgehobener Bedeutung sind die Odds-Ratio (OR) mit

$$OR = \frac{odds(x+1)}{odds(x)} = e^b, \tag{3.33}$$

wobei b wie bereits oben für den jeweiligen Koeffizienten der unabhängigen Variable steht. Dieser Ausdruck lässt sich insofern interpretieren, dass sich die Odds mit der Zunahme einer Einheit von x und damit die Chance für den Ereigniseintritt um den Faktor e^b erhöhen, weshalb die OR auch unter der Bezeichnung Effekt-Koeffizient zu finden sind. In diesem Zusammenhang ist es für die Interpretation wichtig, dass die Skalierung der unabhängigen Variablen beachtet wird. Bei einem negativen Regressionskoeffizienten wird der Wert für den Faktor kleiner eins. Dies muss so interpretiert werden, dass eine Verringerung der Odds um diesen Faktor stattfindet bei einer zusätzlichen Einheit von x. Nimmt der Koeffizient den Wert null an, so wird das OR eins und folglich hat diese Variable keinen Einfluss auf die abhängige Variable.

Logit

Eine weitere Möglichkeit für die Interpretation von Regressionskoeffizienten sind die logarithmierten Odds (kurz: Logit), die sich über die Formel

$$logit(p) = ln(odds(p)) = ln\left(\frac{p}{1-p}\right) = z(x) \tag{3.34}$$

berechnen lassen. Durch die Logitbildung wird der Wertebereich auf $[-\infty; \infty]$ ausgedehnt. Analog zur OR-Formel 3.33 erhöht sich der Wert des Logits um b, wenn sich x um eine Einheit erhöht.

Zusammengefasst lässt sich die Interpretation der Koeffizienten aufgeteilt auf die einzelnen Verfahren folgendermaßen darstellen:

	Erhöhung von x um eine Einheit	
	$b > 0$	$b < 0$
odds	Erhöhung um Faktor e^b	Verminderung um Faktor $e^{-\|b\|}$
logit	Erhöhung um den Betrag b e^b	Verminderung um den Betrag $\|b\|$
Odds-Ratio	Es gilt: $e^b > 1$	Es gilt: $e^b < 1$

Tabelle 3.3: Wirkungen des Zuwachses von x um eine Einheit unter Beachtung des Vorzeichens des Regressionskoeffizientens (e.D. in Anlehnung an Backhaus et al. (2016, S. 313))

Relatives Risiko (RR)

Die bereits beschriebenen Möglichkeiten zur Interpretation der Regression setzen grundsätzlich metrische Daten voraus, die sich um eine Einheit erhöhen lassen. Für binäre Variablen ist dieses Vorgehen wenig aussagekräftig, sodass das RR die geeignetere Darstellungsform ist. Hierfür wird die Wahrscheinlichkeit, wenn die unabhängige Variable den Wert eins annimmt, mit der Wahrscheinlichkeit für den Fall, dass die Variable null ist, ins Verhältnis gesetzt. Die Werte für die unterschiedlichen Wahrscheinlichkeiten bei einer einzelnen Beobachtung errechnen sich dadurch, dass die betrachtete Variable in ihren Ausprägungen variiert wird und die anderen unabhängigen Variablen konstant bleiben, wofür die Formel 3.10 Verwendung findet. Dies lässt sich exemplarisch bei zwei Gruppen für eine abhängige Variable darstellen durch:

$$RR_0 = \frac{p_0}{p_1} \text{ bzw. } RR_1 = \frac{p_1}{p_0}. \tag{3.35}$$

4 Daten

4.1 Datensatz

Der dieser Bachelorarbeit zugrunde liegende Datensatz stammt aus den Befragungsergebnissen der im Jahr 2013 zum zehnten Mal durchgeführten Verkehrserhebung 'Mobilität in Städten - SrV' und stellt die aktuellste Ausgabe dieser Befragungsreihe dar. Diese Form der Erhebung wurde 1972 als SrV ins Leben gerufen und wird seitdem in einem regelmäßigen Turnus durchgeführt (vgl. Ahrens et al. 2014, S. 1). Diese Arbeit stützt sich auf die Datensätze aus den 25 großen SrV-Städten, die sowohl ost- als auch westdeutsche Großstädte umfassen. Eine Übersicht der verwendeten Städte befindet sich im Anhang in Tabelle A.1. Aus den vorliegenden Datensätzen sind keine Rückschlüsse möglich, in welcher Stadt sich der jeweilige Haushalt befindet, sodass keine Aussagen zu raumspezifischen Besonderheiten getroffen werden können.

Die umfangreiche Datenerhebung verfolgt verschiedene Zwecke. Zum einen soll das Verkehrsverhalten der Bevölkerung erhoben und analysiert werden (vgl. Ahrens et al. 2014, S. 1). Zum anderen steht der Erkenntnisgewinn und die Sammlung von Basisinformationen im Hinblick auf die integrierte Verkehrsentwicklung im Vordergrund (vgl. ebd.). Zudem dienen die Daten als Grundlage für statistische Auswertungen.

Erhebungsmethodik

Für die Datenerhebung wurden vorab einige methodische Festlegungen getroffen, die im Folgenden kurz erläutert werden. Die Befragung der Personen in den Haushalten erfolgte mittels telefonischer oder schriftlicher Verfahren, wobei die Befragten zwischen den beiden Formen wählen konnten (vgl. Ahrens et al. 2014, S. 22). Der Erhebungszeitraum umfasste das gesamte Jahr 2013, wobei als Stichtage die mittleren Werktage (Dienstag, Mittwoch, Donnerstag) definiert wurden (vgl. Ahrens et al. 2015, S. 3).

4.2 Aufbereitung der Daten

Die Daten dieser Arbeit basieren auf drei separaten Datensätzen, die verschiedene Informationen auf Haushalts-, Personen- und Wegeebene enthalten. Diese Multilevelstruktur kann für eine Gesamtanalyse nicht herangezogen werden. Für die Auswertung müssen die Daten dementsprechend auf ein einheitliches Niveau gebracht werden. Die Variablen wurden aus sachlogischen Gründen einer Vorauswahl unterworfen und anschließend beurteilt, ob die Daten im Hinblick auf die tiefergehende Analyse in die weitere Betrachtung aufgenommen werden sollen. In der ursprünglichen Datenversion liegen Informationen über 23.972 Haushalte mit insgesamt 54.579 Personen vor, die an der Umfrage teilgenommen haben. Gleichzeitig wurden in der Erhebung Informationen über 178.337 Wege erfasst. Diese große Stichprobe lässt auf gute Ergebnisse hoffen,

die möglichst realitätsnah die Grundgesamtheit der Wohnbevölkerung in den Städten beschreibt. Der Pkw-Besitz wird auf Haushaltsebene erfasst, weshalb die unterschiedlichen Datensätze auf dieser Ebene zusammengefasst werden müssen.

4.2.1 Zusammenhänge in der Multilevelstruktur

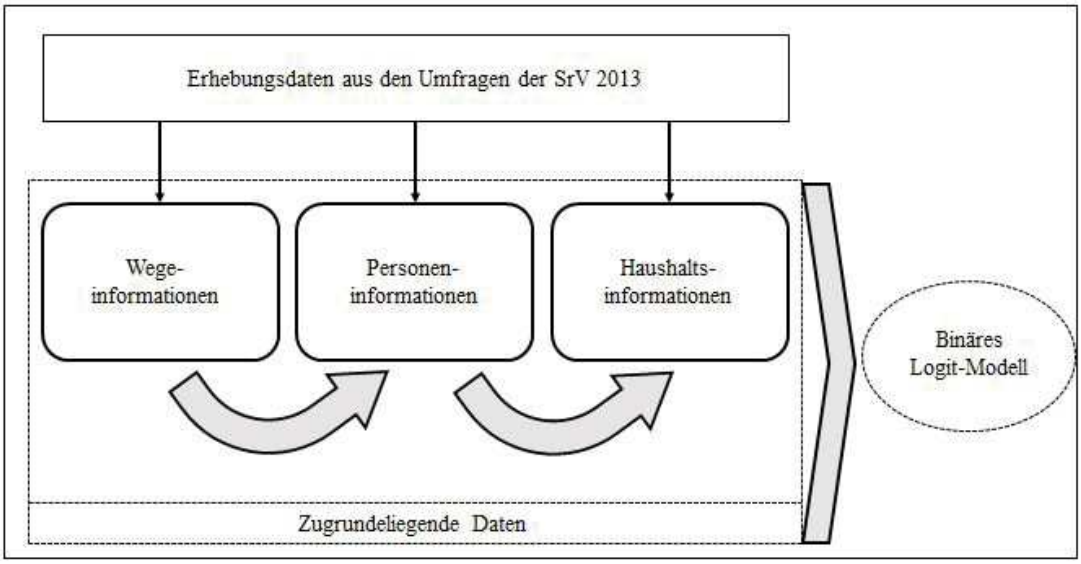

Abbildung 4.1: Multilevelstruktur der verwendeten Daten (e.D.)

Wie bereits erläutert, ist es für die Analyse zielführend, wenn die Daten auf Ebene der einzelnen Haushalte vorliegen. Hierzu mussten Überlegungen angestellt werden, unter welchen Aspekten die Komprimierung der verschiedenen Datensätze durchgeführt werden kann. Ziel ist dabei, die Charakteristika der Haushalte möglichst nah an den einzelnen Personen zu erhalten. Die Haushaltsdatei stellt dabei die Grundversion dar, in der die verschiedenen Datensätze vereinigt werden sollen. In der Personendatei sind Informationen über die Haushaltsmitglieder enthalten mit bis zu zehn Personen pro Haushalt. Die verschiedenen Wege der Haushaltsmitglieder sind im dritten Datensatz aufgeführt. Hierin sind Informationen über die Wegerelationen der einzelnen Haushaltsmitglieder verfügbar. In der vorliegenden Stichprobe sind bis zu 21 Wege pro Person an einem Stichtag aufgeführt. Wie in Abbildung 4.1 ersichtlich ist, wurden in einem ersten Schritt die Wegedaten auf die Personenebene komprimiert und im nächsten Schritt mit dieser zusammengeführt. Im Anschluss daran wurden die Personendaten zusammengefasst und dem jeweiligen Haushalt zugeordnet, um die Multilevelstruktur aufzuheben. Dieser Datensatz wurde für die weitere Analyse und für die Modellschätzung verwendet.

4.2.2 Wegedaten

Aufgrund der im Abschnitt 4.2.1 beschriebenen Multilevelstruktur sind in einem ersten Schritt die Wegedaten aufzubereiten. Im Vorfeld der Komprimierung sind Überlegungen notwendig, welche Informationen über die Wege einen Einfluss auf den Pkw-Besitz der Haushalte haben. Neben den Ergebnissen der Befragung erscheint es zielführend, weitere Variablen zu generie-

ren, die einen Einfluss auf den Besitz von Fahrzeugen eines Haushaltes haben könnten. Unter der Annahme, dass das Angebot der öffentlichen Verkehrsmittel in der Schwachlastzeit weniger ausgeprägt ist, liegt es nahe zu berücksichtigen, ob ein Weg besonders in den Relationen Wohnen-Arbeiten oder Arbeiten-Wohnen in diesem Zeitraum stattgefunden hat. In dieser Arbeit wird die Schwachlastzeit für den Zeitraum von 20 Uhr bis 05:59 Uhr definiert. Darüber hinaus könnte die Tatsache, dass in den oben genannten Relationen ein weiteres Ziel angesteuert wird, aufgrund der höheren Flexibilität, eher auf ein Auto hinweisen. So wurden alle Wege, die in den Relationen Wohnen-Arbeiten und Arbeiten-Wohnen ein weiteres Ziel innerhalb eines Zeitraumes von zwei Stunden ansteuern, über eine binäre Variable berücksichtigt. Am Ende der Datenselektion liegen die Daten für jede Person eines Haushalts vor, die Informationen über einen Weg angegeben hat und deren Weg gültig war. Bei einem gültigen Weg sind Angaben über Dauer und Länge vorhanden. Zudem betrug die Länge eines einzelnen Weges weniger als 100 Kilometer am Stichtag. In der Tabelle 4.1 sind die in Frage kommenden Variablen, die Skalierung und die genutzte Komprimierungsfunktion zusammengefasst.

Variable	**Skalierung**	**Komprimierungsfunktion**
Zeit	nominal	Maximum
Zwischenstation Relation WA	nominal	Maximum
Zwischenstation Relation AW	nominal	Maximum
Begleitung einer Person	nominal	Maximum
Anzahl der begleiteten Personen	kardinal	Maximum
Wegelänge	kardinal	Summe

Tabelle 4.1: Variablen aus der Wegeebene (e.D.)

4.2.3 Personendaten

In der Personendatei sind hauptsächlich sozioökonomische Information und Attribute über das Verkehrsverhalten, wie beispielsweise die Nutzung einer Dauerkarte oder der Besitz eines Führerscheins aller im Haushalt lebenden Personen, enthalten. Sind zu einer Person Wegeinformationen aus dem vorherigen Schritt vorhanden, werden diese den jeweiligen Personen zugeordnet. Der Besitz eines Pkws ist für einen längeren Zeitraum ausgelegt, sodass für das Modell langfristig angelegte Informationen von Bedeutung sind. Aus diesem Grund sind in der Personendatei alle Personeninformationen und deren Wege eliminiert worden, für die der Stichtag kein ‚normaler' Tag war (vgl. Ahrens et al. 2014, Anhang II-S. 11). Dieser ist dadurch charakterisiert, dass die Abläufe am Stichtag den anderen Tagen des gleichen Wochentags ähneln. Vor dem Hintergrund, dass auch Daten von Minderjährigen enthalten sind, muss ein weiterer Schritt in der Datenaufbereitung erfolgen. Es ist davon auszugehen, dass die sozioökonomischen Informationen der Kinder wie zum Beispiel Schulausbildung oder der Besitz von Dauerkarten, die oftmals kostenfrei für den Schulweg zur Verfügung stehen, keinen Einfluss auf die Autoausstattung der Haushalte haben. Somit wurden diese Daten entfernt und nur Informationen von volljährigen Personen berücksichtigt. Auch bei der Berechnung des Durchschnittsalters der Personen in den Haushalten sind Minderjährige herausgefallen. Dagegen ist die Betreuung von Kindern eine wichtige Variable in der Modellierung, die über eine Dummy-Variable in das Modell mitaufgenommen

wurde (siehe Kapitel 5.5). Eine weitere Besonderheit der Multilevelstruktur auf Haushaltsebene stellt die Geschlechterverteilung dar. Mit der Unterscheidung in Ein- und Mehrpersonenhaushalte kann es vorkommen, dass es kein dominierendes Geschlecht in den klassischen Ausprägungen gibt. Um dies aufzugreifen, wurde für die Variable Geschlecht eine dritte Ausprägung ‚gemischt' eingeführt. Hierzu wurde der Durchschnitt über alle Personen gebildet und dem Wert 0,5 die Ausprägung ‚gemischt' zugeteilt. Tabelle 4.2 enthält die für die weitere Analyse verwendeten Daten.

Variable	**Skalierung**	**Komprimierungsfunktion**
Höchster Schulabschluss der Volljährigen	kategorial	Maximum
Höchste Berufsausbildung der Volljährigen	kategorial	Maximum
Geschlecht	kategorial	Durchschnitt
Pkw Führerschein	nominal	Maximum
Sonstiger Führerschein	nominal	Maximum
Dauerkarte bei Volljährigen	nominal	Maximum
Nutzung Bikesharing von Volljährigen	nominal	Maximum
Nutzung Carsharing von Volljährigen	nominal	Maximum
Erwerbstätigkeit	nominal	Minimum
Einschränkung	nominal	Maximum
Wegeanzahl	kardinal	Summe
Alter der Volljährigen	kardinal	Durchschnitt

Tabelle 4.2: Variablen aus der Personenebene (e.D.)

4.2.4 Haushaltsdatei

Zielebene und Ausgangsgröße für das Modell ist die Haushaltsebene. Diese Datei umfasst Informationen über 23.971 Haushalte. Hierunter fällt unter anderem das klassierte Haushaltseinkommen, die Personenanzahl oder die Entfernung zur nächstgelegenen Haltestelle des ÖVs. Gleichzeitig ist in diesem Datensatz die abhängige Variable ‚Anzahl der Pkw' für das Modell enthalten. Aus diesem Datensatz wurden Informationen von Haushalten entfernt, die bei der Anzahl an Pkws keine plausiblen Angaben machten und somit keiner Kategorie zugeordnet werden können. Daraufhin wurden die Ergebnisse aus den vorangegangenen Schritten in die Datei überführt. Die weiteren in das Modell aufgenommenen Variablen sind in Tabelle 4.3 aufgeführt.

Variable	**Skalierung**
Anzahl Personen	kardinal
Anzahl der Pkw	kardinal
Anzahl motorisierter Zweiräder	kardinal
Anzahl unmotorisierter Zweiräder	kardinal
Kürzeste Entfernung zur nächste ÖV-Haltestelle	kardinal
Einkommensklassen	pseudokardinal
Verfügbarkeit Dienstwagen	nominal

Tabelle 4.3: Variablen aus der Haushaltsebene (e.D.)

4.3 Datengrundlage

Ein großer Nachteil der hier dargestellten Multilevelstruktur ist, dass einige Informationen über die einzelnen Personen verloren gehen. Die Aggregation der Daten auf Haushaltsebene birgt zudem die Gefahr, dass mögliche Korrelationen innerhalb der metrischen Daten das Ergebnis beeinträchtigen können. So liegt es nahe, dass Haushalte mit einer größeren Anzahl an Personen mehr Wege zurücklegen oder eine größere Anzahl an Alternativen, wie etwa motorisierte Zweiräder, zur Verfügung stehen. Von einer Aggregation dieser metrischen Variablen mittels Durchschnittsbildung wurde aus dem Grund abgesehen, dass eine Aggregation der Variablen auf Haushaltsebene nicht möglich ist. Zudem sollen einheitliche Maßstäbe für alle Variablen angesetzt werden. Darüber hinaus würde eine zu starke Verzerrung des Gesamtbildes stattfinden, wenn Haushalte mit vielen Personen nur wenige Wege vorweisen, weil eine Person einen längeren Weg zurücklegen muss, die restlichen aber nur kurze Fußwege aufweisen.

Insgesamt liegen dem Modell 23.950 Datensätze zugrunde, darunter 11.986 vollständige Informationssätze. Die hohe Anzahl an fehlerhaften Datensätzen ist dadurch zu begründen, dass durch mangelnde Informationen aus der Personen- und Wegeebene oder infolge von Eliminierungen nicht für alle Haushalte Ergebnisse ermittelt werden konnten. Als Wert für das Einkommen wurde die Mitte der Einkommensklassen aus den Ursprungsdaten für die weitere Betrachtung verwendet. Von den 23.950 Haushalten besitzen 18.777 Haushalte mindestens einen Pkw. Demzufolge setzt sich der Datensatz aus 21,6 % an Haushalten ohne Pkw und 78,4 % an Haushalten mit einem oder mehreren Pkw zusammen. Daraus resultieren unterschiedliche Gruppengrößen.

Die Aufbereitung des Datensatzes erfolgt unter Anwendung des Tabellenkalkulationsprogramms Excel, wobei aufgrund der großen Datenmenge automatisierte Aggregationsverfahren in Form von Visual Basic for Applications (VBA) genutzt wurden.

5 Deskriptive Analyse

5.1 Vorgehen

Nach der theoretischen und logischen Auswahl sowie der Aufbereitung der Variablen soll nun eine weitergehende deskriptive Betrachtung der Daten erfolgen. Damit soll der Einfluss dieser auf das Modell beurteilt und Variablen mit schwacher Aussagekraft eliminiert werden. Gleichzeitig soll eine Untersuchung stattfinden, ob es bereits anhand der deskriptiven Analyse Anhaltspunkte dafür gibt, welche Variablen für eine Zuordnung zu den Gruppen ‚Kein Pkw' oder ‚1 oder mehr Pkw' geeignet sind. Hierbei werden die Variablen aufgrund ihrer Skalierung separat betrachtet. In einem ersten Schritt werden die kardinal skalierten Variablen mithilfe der relativen Häufigkeitsverteilung beurteilt und - wenn möglich - weitere Rückschlüsse aus einem Boxplot gezogen. Der Zusammenhang zwischen den kardinalen Variablen wird mithilfe einer Korrelationsmatrix beurteilt. Die nominal und kategorial skalierten Variablen werden mithilfe des in 3.1.2 vorgestellten korrigierten Kontingenzkoeffizienten nach Karl Pearson beurteilt. Die fehlenden Werte bei den Variablen werden in der deskriptiven Analyse nicht berücksichtigt.

5.2 Streu- und Lagemaße für kardinal skalierte und klassierte Variablen

Im folgenden Abschnitt werden die Streu- und Lagemaße der kardinal skalierten sowie der klassierten Variablen genauer beschrieben. Anhand dieser wird entschieden, ob eine Aufnahme in das Modell zielführend ist. Die Untersuchung der Variablen findet unter dem Gesichtspunkt statt, ob ein Einfluss auf den Pkw-Besitz eines Haushalts durch Vorliegen dieser Eigenschaft zu erwarten ist. Eine Übersicht über alle betrachteten metrischen Variablen ist im Anhang in der Tabelle A.2 zu finden.

5.2.1 Alternativenspezifische Variablen

Merkmalsvariable ‚Wegelänge'

Die Merkmalsvariable ‚Wegelänge' enthält die summierten Längen der zurückgelegten Wege aller im Haushalt lebenden Personen in Kilometern. Für die Auswertung wurden die einzelnen Angaben in Klassen mit einer Klassenbreite von jeweils zehn Kilometer eingeteilt und alle Werte, die mehr als 100 Kilometer Wegelänge aufweisen, zu einer Kategorie zusammengefasst. Im Boxplot sind hingegen die einzelnen Werte verarbeitet, da ein Boxplot für die Darstellung einzelner metrischer Werte konstruiert ist.

Wie aus den relativen Häufigkeiten in Abbildung 5.1 ersichtlich ist, bewegen sich Personen in Haushalten ohne Pkw eher in niedrigeren Distanzklassen. So weisen Haushalte ohne Auto in

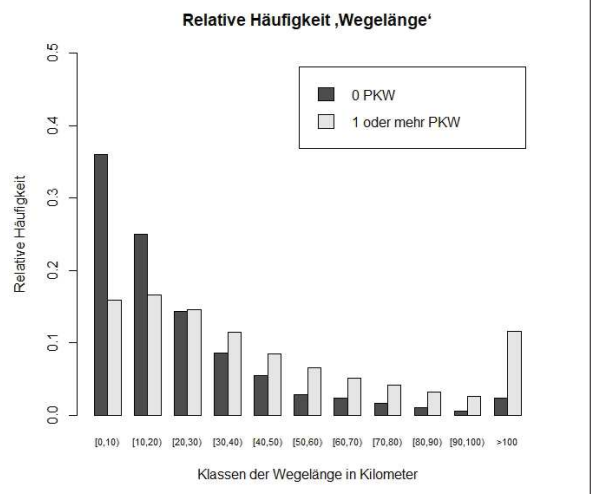

Abbildung 5.1: Relative Häufigkeit ‚Wegelänge‘ (e.D.)

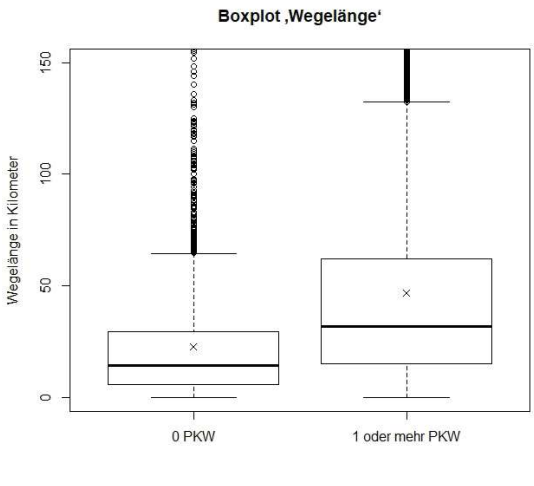

Abbildung 5.2: Boxplot der Variable ‚Wegelänge‘ (e.D.)

75 % der Fälle nur eine Wegelänge bis zu 30 Kilometern auf, während der Anteil der Haushalte mit einem oder mehreren Pkw im gleichen Distanzbereich nur bei 48 % liegt. Mit zunehmender täglicher Wegelänge nimmt der Anteil der Haushalte ohne Pkw im Vergleich zu Pkw-Haushalten weiter ab. Nur 2,3 % aller Wege in 0-Pkw-Haushalten haben eine Länge größer als 100 Kilometer. Im Kontrast dazu weist die andere Gruppe einen Anteil von über 11 % auf. Während der Anteil von 0-Pkw-Haushalten in der Distanzklasse bis zehn Kilometer mehr als doppelt so hoch ist als der von Haushalten mit mindestens einen Pkw, dreht sich dieses Verhältnis ab einer Länge von 30 Kilometern. Damit ist der Anteil der Haushalte mit Pkw ab einem Distanzbereich von 60 Kilometern doppelt so hoch wie bei Haushalten ohne Pkw. Dies entspricht auch der Annahme, dass Wege mit längeren Distanzen eher mit Autos als mit öffentlichen Verkehrsmitteln oder Fahrrädern zurückgelegt werden.

Wird der Boxplot in Abbildung 5.2 betrachtet, so fällt auf, dass sich die Mittelwerte zwischen beiden Gruppen stark unterscheiden. So liegt der Mittelwert bei der Gruppe ohne Pkw mit 22,73 Kilometern wesentlich niedriger als bei der Gruppe mit Pkw, die eine mittlere Wegelänge von 46,91 Kilometern aufweist. Die Mittelwerte liegen zudem deutlich oberhalb der Mediane. Sowohl der Quartilsabstand als auch der Abstand zwischen den Whiskern der beiden Gruppen lässt darauf schließen, dass die Streuung bei Haushalten, die im Besitz eines Pkws sind, wesentlich größer ist. Beide Gruppen weisen den Beginn des ersten Whiskers bei null Kilometern auf. Daraus lässt sich schließen, dass es auch einige Haushalte in der Stichprobe gibt, die keinen täglichen Weg aufweisen. Beide Gruppen weisen eine hohe Anzahl an Ausreißern auf, wobei die Ausreißer mit zunehmender Wegelänge bei Haushalten ohne Pkw deutlich abnehmen. Zudem sind sie rechtsschief. Der gesamte Boxplot der Gruppe mit Pkw ist breiter und in höheren Distanzbereichen situiert. Dies weist auf größere Distanzen in Haushalten mit Pkws hin.

Aus der Analyse der Streu- und Lagemaße lässt sich folgern, dass Haushalte mit Pkw wesentlich längere Distanzen zurücklegen. Daraus kann abgeleitet werden, dass diese Variable eine Trennung zwischen den Gruppen ermöglicht. Folglich scheint es zweckmäßig, diese Variable in das Modell aufzunehmen.

Merkmalsvariable ‚Anzahl der Wege'

Neben der Wegelänge lässt sich auch die Vermutung anstellen, dass je mehr Wege ein Haushalt pro Tag zurücklegen muss, desto eher befindet sich dieser im Besitz eines Pkws. Dies könnte dadurch begründet werden, dass ein Auto eine größere räumliche Flexibilität und Unabhängigkeit ermöglicht.

Die relativen Häufigkeiten in Abbildung A.1 im Anhang zeigen, dass der Anteil von Haushalten ohne Pkw eher im Bereich einer niedrigeren Wegeanzahl höher ist, als der von Pkw-Haushalten. So legen etwa 60 % aller Haushalte ohne Pkw täglich nur bis zu vier Wege zurück. Den größten Anteil weisen dabei zwei Wege pro Haushalt (0,20) auf. Anders verhält es sich bei den Haushalten mit Pkw-Besitz. Deren Anteile verteilen sich relativ gleichmäßig bis zu einer Wegezahl von neun. Dabei bewegt sich der relative Anteil überwiegend bei jeweils etwa 10 %. Den höchsten Anteil nehmen Haushalte mit Pkw-Besitz und mehr als neun Wege pro Tag ein. Folglich legen ein Drittel aller Haushalte mit Pkw täglich mehr als neun Wege zurück, während dieser Anteil bei Haushalten ohne Pkw wesentlich geringer ist.

Der Boxplot in Abbildung A.2 im Anhang bestätigt diese These. So beträgt der Mittelwert bei Haushalten ohne Pkw 4,99 Wege und bei Haushalten mit Pkw 7,46 Wege. Dieses Ergebnis spricht deutlich dafür, dass der Pkw-Besitz mit einer höheren Anzahl von Wegen in Verbindung steht. Auch der Median von Haushalten ohne Pkw liegt deutlich unter dem der Pkw-Haushalte. Werden die Quartilsabstände betrachtet, so ist dieser bei den Haushalten mit Pkw größer, ebenso wie der Abstand der Whisker. Dieses lässt auf eine größere Streuung schließen. In den beiden Boxplots ist Eigenschaft der Rechtsschiefe erkennbar, wobei dieser Effekt bei den Pkw-Haushalten stärker auftritt. Beide Gruppen weisen Ausreißer auf, die plausibel mit unterschiedlichen Gegebenheiten, wie einer hohen Anzahl täglicher Wege bei der Kinderbetreuung oder im Arbeitsalltag, begründbar sind.

In der Folge lässt sich anhand der deskriptiven Untersuchung sagen, dass Haushalte mit mehr Wegen eher zum Pkw-Besitz tendieren und damit eine Trennung erwarten lassen und förderlich für das Modell sind.

Merkmalsvariable ‚Kürzester Weg zur nächsten ÖV-Haltestelle'

Alternativen zum eigenen Auto wirken sich ebenfalls auf den Pkw-Besitz aus. So wird in dieser Arbeit der kürzeste Weg von der Wohnung zur nächsten ÖV-Haltestelle in Minuten betrachtet. In den untersuchten Städten ist aufgrund ihrer Größe davon auszugehen, dass überwiegend ein gutes Angebot im öffentlichen Personenverkehr besteht und die Haushalte zwischen verschiedenen Verkehrsmitteln (U-Bahn, Straßenbahn, Bus, etc.) wählen können. Trotzdem gibt es Haushalte, die über keine ÖV-Anbindung verfügen. Um dies abbilden zu können, wird diesen Haushalten eine Entfernung von 60 Minuten zur nächsten ÖV-Haltestelle zugewiesen, was einen Widerstand darstellt, der in der Realität dazuführt, dass der ÖV nicht als Alternative angesehen wird. Um eine übersichtliche Darstellung zu gewährleisten, wurde eine Klasseneinteilung für die Entfernung zur nächsten Haltestelle in Minuten vorgenommen.

Sowohl bei den relativen Häufigkeiten in Abbildung A.3 im Anhang als auch im Boxplot in Abbildung A.4 im Anhang lassen sich keine bedeutenden Unterschiede zwischen den Haushalten ohne und mit Pkw erkennen. Bei den relativen Häufigkeiten überwiegt bis zu einer Entfernung von sechs Minuten zur nächsten Haltestelle der Anteil von Haushalten ohne Pkw, wobei der

Unterschied nur einem Anteil von etwa 0,01 entspricht. Daraus kann gefolgert werden, dass Haushalte mit einer kurzen Entfernung zur nächsten Haltestelle den ÖV leicht präferieren. Ab Entfernungen von sieben Minuten bis zur nächsten Haltestelle dreht sich dieses Verhältnis und der Anteil von Haushalten mit Pkw überwiegt. Jedoch ist auch hier der Unterschied minimal. Insgesamt liegen in beiden Gruppen über 80 % aller Haushalte in den niedrigen Entfernungsbereichen bis sechs Minuten zur nächsten Haltestelle. Zurückzuführen könnte dieses Ergebnis auf die urbane Struktur der Datengrundlage sein und damit für eine gute Versorgung mit ÖV-Leistungen sprechen.

Dieser Eindruck bestätigt sich im Boxplot in Abbildung A.4 im Anhang, wobei sowohl Quartilsabstand, Abstand zwischen den Whiskern, die Quartile und der Median in beiden Gruppen identisch ist. Dies kann darauf zurückgeführt werden, dass ein Großteil der Haushalte nahe einer Haltestelle liegt. Einzig der Mittelwert lässt einen minimalen Unterschied erkennen, wobei dieser bei Haushalten ohne Pkw mit 4,39 Minuten etwas geringer ausfällt als bei Haushalten mit Pkw-Besitz mit 4,57 Minuten. Dieser Unterschied könnte auf die höhere Anzahl an Ausreißern bei Haushalten mit Pkw-Besitz zurückzuführen sein.

Folglich lässt sich eine relativ geringe Trennwirkung bei der Entfernung zur nächsten Haltestelle identifizieren. Eine weitere Betrachtung dieser doch sehr aussagekräftigen Variable wird im Modell trotzdem angestrebt, da sich eine Trennwirkung gemäß den Annahmen erwarten lässt.

Merkmalsvariablen ‚Anzahl von motorisierten/unmotorisierten Zweirädern'

Eine weitere Alternative zum Pkw stellen Zweiräder dar. In dieser Arbeit wird zwischen motorisierten Zweirädern (Motorrad) und unmotorisierten Zweirädern (Fahrrad, E-Bike) unterschieden sowie auf die Anzahl dieser Verkehrsmittel pro Haushalt eingegangen. Dabei wird die Anzahl der Zweiräder in den Haushalten summiert ausgewiesen.

Merkmalsvariable ‚Anzahl motorisierter Zweiräder'

Die relativen Häufigkeiten in Abbildung A.6 im Anhang zeigen, dass ein sehr großer Anteil von Haushalten nicht im Besitz eines motorisierten Zweirades ist. Bei den Haushalten ohne Pkw besitzen 96 % kein motorisiertes Zweirad und bei den Haushalten, die über mindestens einen Pkw verfügen, liegt der Anteil noch bei 88 %. Knapp 10 % der Haushalte mit Pkw verfügen über ein motorisiertes Zweirad, während nur knapp 4 % der Pkw-losen Haushalte ein Motorrad oder ähnliches Zweirad besitzen. Daraus lässt sich folgern, dass motorisierte Zweiräder keine Alternative zum Pkw-Besitz darstellen. Daraus lässt sich auch eine Affinität in Richtung Motorisierung erahnen, da motorisierte Haushalte eher auch über motorisierte Zweiräder verfügen.

Eine Darstellung als Boxplot ist aufgrund der Konzentration der Werte im niedrigen Bereich nicht sinnvoll. Im Hinblick auf die Trennwirkung anhand dieser Variablen lässt sich aussagen, dass der Besitz eines Motorrades ein Hinweis auf den Pkw-Besitz ist. Demzufolge findet diese Variable im Modell Berücksichtigung.

Merkmalsvariable ‚Anzahl unmotorisierter Zweiräder'

Anhand der relativen Häufigkeiten der unmotorisierten Zweiräder in Abbildung A.7 im Anhang fällt auf, dass Haushalte ohne Pkw eher dazu tendieren kein oder nur wenige Fahrräder zu besitzen. Über ein Drittel der Haushalte ohne Pkw verfügt auch über kein Fahrrad, was die

Vermutung widerlegt, dass unmotorisierte Zweiräder bedingungslos als Alternative angesehen werden. Zudem besitzen Haushalte mit mehr als einem Fahrrad eher mindestens einen Pkw. Dies lässt sich möglicherweise auf die Anzahl der Personen im Haushalt zurückführen, wobei diese eher zu einem Fahrrad pro Person tendieren und zudem, wie oben dargestellt, eher im Besitz eines Pkws sind. Aufgrund dessen lässt sich keine eindeutige Trennwirkung aus dieser Variablen erahnen, sodass diese im Modell keine Betrachtung findet.

5.2.2 Sozioökonomische Variablen

Merkmalsvariable ‚Anzahl der Personen'

Aus sachlogischen Überlegungen liegt es nahe, dass ein Haushalt mit mehreren Personen tendenziell ein Auto hat, da sich der Besitz aus ökonomischen Gründen rechnet und eine komfortablere Beförderung von gewöhnlich bis zu fünf Personen gleichzeitig ermöglicht.

Bei der Merkmalsvariable ‚Anzahl der Personen' fällt bei den relativen Häufigkeiten in Abbildung A.11 auf, dass über 65 % aller Haushalte beider Gruppen aus Ein- und Zweipersonenhaushalten bestehen. In den Haushalten ohne Pkw leben in 59 % der betrachteten Haushalte nur eine Person und in 27 % zwei Personen. Haushalte mit Pkw haben überwiegend zwei oder mehr Haushaltsmitglieder. Somit wird die Vermutung bestätigt, dass Haushalte mit einer höheren Personenanzahl eher im Besitz eines Pkws sind. Haushalte mit mehr als fünf Haushaltsmitglieder sind in der Stichprobe in beiden Gruppen kaum vertreten.

Aus den Ergebnissen des Boxplots in Abbildung A.12 im Anhang wird das Bild der relativen Häufigkeiten bestätigt. Der Median bei Pkw-losen Haushalten befindet sich bei Einpersonenhaushalten, während der Median in Haushalten mit Pkw-Besitz bei zwei Personen pro Haushalt liegt. Beide Mediane entsprechen den 1. Quartilen der Stichprobe und sind daher stark rechtsschief. Vergleicht man die Mittelwerte in den beiden Gruppen, so liegt der Mittelwert der Pkw-Haushalte (2,45) deutlich über dem der Haushalte ohne Pkw (1,66), wodurch der Eindruck bestätigt wird. Die Streuung in beiden Gruppen fällt eher gering aus, sodass sowohl der Quartilsabstand, als auch die Länge der Whisker nur jeweils eine Person umfassen. Bei Pkw-losen Haushalten existieren keine Whisker unterhalb des Medians, was insofern nachvollziehbar ist, dass es keine Haushalte ohne Personen gibt. Beide Gruppen weisen viele Ausreißer nach oben hin auf. Dies erscheint unter dem Aspekt plausibel, dass es in städtisch geprägten Regionen auch große Haushalte gibt, die aufgrund des guten Angebots im ÖV oder kurzer Wege komplett auf ein eigenes Fahrzeug verzichten.

Anhand der beiden Abbildungen und der sich daraus ermittelbaren Ergebnisse lässt sich folgern, dass eine Trennung bzw. Einordnung der Haushalte anhand dieser Merkmalsvariable gut möglich ist und Haushalte mit mehreren Personen eher zum Pkw-Besitz tendieren, weshalb die Aufnahme der Variable in das Modell zielführend erscheint.

Merkmalsvariable ‚Einkommen der Haushalte'

Die Merkmalsvariable ‚Einkommen der Haushalte' stellt eine Besonderheit unter den metrischen Variablen dar. Da in der Datengrundlage das monatliche Einkommen in Klassen angegeben ist, wurde aus Gründen der Einfachheit eine pseudokardinale Variable gebildet, die die Klassenmitte der jeweiligen Einkommensklasse widerspiegelt. Einkommen über 5.600 € pro Monat sind in der

höchsten Einkommensklasse aggregiert.

Aus Abbildung A.8 im Anhang mit den relativen Häufigkeiten wird die Vermutung bestätigt, dass mit steigendem Einkommen der Anteil von Haushalten mit Pkw über dem der Haushalte ohne Pkw liegt. So liegen knapp 55 % aller Haushalte ohne Pkw im Einkommensbereich bis etwa 1.200 €. Der Anteil von Haushalten mit Pkw-Besitz ist in diesem Bereich wesentlich geringer (0,14). Ab der Klasse mit der Einkommensmitte von 2.300 € überwiegt der Anteil von Haushalten mit Pkw deutlich gegenüber dem Anteil Pkw-loser Haushalte, so ist der Anteil bereits in der Gruppe mit der Klassenmitte von 2.300 € doppelt so groß (0 Pkw: 0,05 | $\geq$ 1 Pkw: 0,11). In der Klasse mit den höchsten Einkommen überwiegt deutlich der Anteil von Haushalten mit Pkw-Besitz.

Daraus lässt sich schließen, dass das Einkommen eine große Wirkung auf den Pkw-Besitz entfaltet und mit zunehmenden Einkommen die Ausstattung mit Autos ebenfalls zunimmt. Aus diesem Grund ist eine Trennwirkung wahrscheinlich und eine Aufnahme in das Modell lässt eine Verbesserung der Qualität erwarten.

Merkmalsvariable ‚Durchschnittsalter der Haushalte'

Die sozioökonomische Merkmalsvariable ‚Durchschnittsalter der Haushalte' wird aus dem Alter der volljährigen Personen in einem Haushalt gebildet. Die Annahme der Volljährigkeit wird damit begründet, dass Personen unter 18 Jahren selbst wenig bis kein Mitspracherecht über den Pkw-Besitz haben und so bei der Modellierung nicht berücksichtigt werden.

Aus der Abbildung A.9 im Anhang mit den relativen Häufigkeiten des Durchschnittsalters der volljährigen Haushaltsmitglieder lässt sich erahnen, dass der Pkw-Besitz überwiegend in Haushalten dominiert, dessen Durchschnittsalter in der Klasse zwischen 30 und 64 Jahren liegt (0,51 gegenüber 0,72). Insgesamt haben über die Hälfte der Haushalte in beiden Gruppen ein Durchschnittsalter von 30 bis 65 Jahren. Zudem ist erkennbar, dass sowohl in der Altersgruppe bis 30 Jahre als auch in der Altersgruppe ab 65 Jahre die Haushalte ohne Pkw dominieren, woraus gefolgert werden kann, dass Haushalte mit jungen Erwachsenen und Senioren eher ohne Pkw agieren. In der Altersgruppe mit dem Durchschnittsalter von 30 bis 64 Jahren dominieren klar Haushalte mit Pkw-Besitz. Dies entspricht ebenfalls der Annahme, dass junge Erwachsene aufgrund der Unabhängigkeit und knapper Mittel eher keinen Pkw besitzen und Rentner freiwillig oder aus anderen Gründen auf einen Pkw verzichten.

Im Gegensatz zur relativen Häufigkeit lassen sich anhand des Boxplots in Abbildung A.10 keine großen Unterschiede zwischen beiden Gruppen ausmachen. Sowohl bei der Gruppe ohne Pkw als auch bei der Gruppe mit Pkw liegt das mittlere Durchschnittsalter der Haushalte etwa bei 50 Jahren. Der Median liegt ebenfalls in beiden Gruppen auf nahezu dem selben Wert (49 Jahre) und befindet sich dazu sehr nah am Mittelwert. Der einzige bedeutende Unterschied, der sich aus dem Boxplot ergibt, ist, dass der Quartilsabstand in Haushalten ohne Pkw größer ist. Daraus kann auf eine größere Streuung geschlossen werden. Die Länge der Whisker ist ebenfalls in beiden Fällen nahezu gleich.

Anhand des Boxplots könnte vermutet werden, dass kein großer Unterschied zwischen den beiden Gruppen besteht. Dies ist darauf zurückzuführen, dass ein großer Teil der Haushalte ein annähernd gleiches Durchschnittsalter aufweist. Die relative Häufigkeit lässt jedoch eher eine gewisse, wenn auch geringe Trennwirkung erwarten, sodass in Haushalten mit durchschnittlich

jüngeren oder älteren Personen eher kein Pkw vorhanden ist. Da das Alter jedoch eine wichtige sozioökonomische Variable für die Beurteilung und Analyse ist, soll eine Berücksichtigung im Modell erfolgen, um deren Wirkung abschätzen zu können.

Merkmalsvariable ‚Anzahl begleiteter Personen'

Neben der Tatsache, ob eine Haushaltsperson im Tagesverlauf eine andere Person begleitet oder nicht, ist auch die Anzahl der Personen von Bedeutung, die begleitet werden. So kann die Begleitung einzelner Personen relativ problemlos auch mit öffentlichen Verkehrsmitteln erfolgen. Erfolgt die Begleitung von zwei oder mehr Personen, ist der Einsatz eines Pkws wahrscheinlicher. Eine derartige Begleitung ist für den Weg zum Kindergarten, zur Schule oder zum Arzt vorstellbar.

Wie aus Abbildung A.5 im Anhang mit den relativen Häufigkeiten der begleiteten Personen hervorgeht, unterscheiden sich die beiden Gruppen nur sehr gering bei den einzelnen Personen. Eine eindeutige Trennung ist nicht zu erkennen. Auch eine Tendenz, dass mit zunehmender Anzahl begleiteter Personen ein Pkw bevorzugt wird, lässt sich nicht erahnen. Folglich lässt sich keine Trennwirkung erahnen, sodass diese Variable im Modell keine Berücksichtigung findet.

5.3 Korrelation zwischen den metrischen Variablen

Zur Aufdeckung von Zusammenhängen zwischen den einzelnen metrischen Variablen dient die Korrelationstabelle, die unter 3.1.2 methodisch beschrieben ist. Gleichzeitig unterstützt diese bei der Aufdeckung von Multikollinearität zwischen den Variablen. Beispielsweise liegt es nahe, dass ein Haushalt mit einer höheren Anzahl an Personen mehr Einkommen erzielt oder eine höhere Anzahl an täglichen Wegen zurücklegt. Um den Zusammenhang und dessen Stärke beurteilen zu können, wird auf Basis der vorgestellten Variablen eine Korrelationstabelle errechnet. Diese Tabelle ist im Anhang unter A.3 abgebildet. Die Werte der Korrelationen zwischen den Variablen bewegen sich im Bereich von -0,313 bis 0,637, was für einen mittleren positiven bzw. negativen Zusammenhang spricht. Auf Grundlage der deskriptiven Statistik wird die Variable ‚Anzahl unmotorisierter Fahrräder' nicht in das Modell aufgenommen, sodass der nächste höhere Wert eine Korrelation zwischen der Wegeanzahl und der Personenanzahl mit 0,562 ist und einen mittleren positiven Zusammenhang ausdrückt. Dies ist unter dem Aspekt nachvollziehbar, dass mehr Personen in der Summe auch mehr Wege zurücklegen. Insgesamt lässt sich aus der Korrelationstabelle kein Wert finden, der eine starke Korrelation angibt, wodurch die Aufnahme der Variablen in das Modell plausibel ist. Der größte Teil der Variablen korreliert in einem schwachen Maße miteinander, sodass diese für eine gute Schätzung des Modells geeignet erscheinen.

5.4 Relative Häufigkeiten kategorialer Variablen

Zusätzlich zu den bereits behandelten und diskutierten Lagemaßen gibt es noch fünf weitere sozioökonomische Variablen, die in Kategorien unterteilt sind und deren Beurteilung anhand der relativen Häufigkeiten erfolgt. Die einzelnen Kategorien gehen jeweils über Dummys in das Modell ein.

5.4.1 Höchste Schulausbildung im Haushalt

Um Aussagen über die Schulbildung und den Zusammenhang mit dem Pkw-Besitz treffen zu können, wurde im Rahmen der Datenaufbereitung der höchste Schulabschluss der volljährigen Haushaltsmitglieder ermittelt. Unterschieden werden bei dieser Variable vier Kategorien, die aus Abbildung A.13 im Anhang ersichtlich sind.

Betrachtet man die relativen Häufigkeiten in Abbildung A.13 im Anhang, fällt auf, dass sowohl bei Haushalten ohne Pkw (0,52) als auch bei den Haushalten mit Pkw (0,58) ein Abitur als höchsten Schulabschluss vorgewiesen werden kann. Insgesamt weisen mehr als die Hälfte aller Haushalte in beiden Gruppen diesen Abschluss auf. Zudem überwiegen in Haushalten mit Abitur oder Realschulabschluss als höchsten Schulabschluss die Haushalte mit Pkw-Besitz. In Haushalten, die als höchsten Schulabschluss einen Hauptschulabschluss aufweisen können, ist der Anteil von Haushalten ohne Pkw mit 20 % höher als der von Pkw-Haushalten mit 10 %. Der Anteil von Haushalten ohne Abschluss ist jeweils sehr gering mit etwa 2,5 %. Ein bedeutender Unterschied zwischen beiden Gruppen ist in dieser Kategorie nicht zu erkennen. Insgesamt lässt sich schlussfolgern, dass je höher der Schulabschluss ist, desto eher besitzen die Haushalte mindestens einen Pkw. Dies entspricht auch der Vermutung, dass mit zunehmenden Schulabschluss ein höheres Einkommen zu erzielen ist und damit der Besitz eines Pkws wahrscheinlicher ist.

5.4.2 Höchste Berufsausbildung im Haushalt

Analog zur Schulausbildung in 5.4.1 wird die höchste Berufsausbildung behandelt.

Ein ähnliches Bild wie bereits bei der Schulausbildung ergibt sich bei der höchsten Berufsausbildung in Abbildung A.14 im Anhang mit den relativen Häufigkeiten. So überwiegt der Anteil von Haushalten mit Pkw-Besitz sowohl in der Kategorie ‚Meister' (0 Pkw: 0,09 | $\geq$ 1 Pkw: 0,13) als auch in der Kategorie ‚Hochschule' (0 Pkw: 0,37 | $\geq$ 1 Pkw: 0,58). Die meisten Haushalte weisen in beiden Gruppen einen Hochschulabschluss auf, knapp gefolgt von Haushalten mit einer Lehre als höchsten Berufsabschluss. In den Kategorien ‚Lehre' und ‚ohne Berufsausbildung' existieren mehr Haushalte ohne Pkw als mit Pkw-Besitz bei einer Differenz von etwa 0,07 zwischen beiden Anteilen. Auch hier wird die Annahme bestätigt, dass mit zunehmender Berufsausbildung der Pkw-Besitz steigt und damit eine Trennwirkung im Modell zu erwarten ist.

5.4.3 Geschlecht

Eine weitere sozioökonomische Variable stellt das Geschlecht dar, dass wie bereits in Kapitel 4.2.3 beschrieben, drei Ausprägungen - weiblich, männlich, gemischt - annehmen kann. Dabei wird einem Haushalt das Attribut ‚weiblich' zugeschrieben, wenn dieses Geschlecht dominiert wie beispielsweise bei einem Haushalt mit alleinerziehender Mutter und einem Sohn oder in einer Wohngemeinschaft mit zwei Frauen und einem Mann. Die Ausprägung ‚gemischt' wird dann zugeteilt, wenn in einem Haushalt die gleiche Zahl an weiblichen und männlichen Personen existieren und keine eindeutige Zuordnung möglich ist.

Wie in Abbildung A.15 im Anhang gezeigt wird, besitzen Haushalte mit überwiegend männlichen Personen nahezu einen identischen Anteil in beiden Kategorien. Beim weiblichen Geschlecht

liegt der Anteil an Haushalten ohne Pkw (0,47) wesentlich höher als bei Haushalten mit Pkw-Besitz (0,29). Dieses Bild dreht sich bei Haushalten mit der Geschlechtsausprägung ‚gemischt', wo Pkw-Haushalte mit 40 % einen wesentlich höheren Anteil einnehmen als Haushalte ohne Pkw mit 21 %. Daraus kann gefolgert werden, dass weiblich geprägte Haushalte eher zur Nutzung von Alternativen zum eigenen Pkw tendieren, während Haushalte mit beiden Geschlechtern eher einen Pkw besitzen. Bei männlich geprägten Haushalten lässt sich keine eindeutige Tendenz erkennen. Anhand dieser Darstellung lässt sich die Vermutung anstellen, dass auch hier eine Trennwirkung erzielt werden kann und das Modell daher mit dieser Dummy-Variable geschätzt werden soll.

5.4.4 Altersklassen

Neben dem in Abschnitt 5.2.2 diskutierten Durchschnittsalter der Volljährigen im Haushalt, ist es von besonderer Bedeutung, die verschiedenen Altersklassen zu berücksichtigen. So ist davon auszugehen, dass der Pkw-Besitz bei Haushalten mit Kindern wahrscheinlicher ist als bei Singlehaushalten, da mehr Wege zurückzulegen sind. Zudem liegt es nahe, dass mit zunehmendem Alter und abnehmender Fahrfähigkeit weniger das Auto und mehr die Alternativen wie der ÖV genutzt werden. Aus diesem Grund ist das Alter der Personen im Haushalt in verschiedene Altersklassen eingeteilt und mithilfe von Dummy-Variablen in das Modell aufgenommen worden. Die Grenzen orientieren sich dabei an klassischen Ereignissen, wie dem Eintritt der Volljährigkeit. Eine weitere Altersklasse wird bei 30 Jahren gebildet, da ab diesem Alter eine gewisse Berufserfahrung und Pläne für eine Familiengründung zu erwarten sind. Auch das Renteneintrittsalter stellt eine weitere Grenze dar. Wichtig ist hier, dass pro Haushalt mehrere Kategorien gleichzeitig vorliegen können. So kann in einem Mehrpersonenhaushalt sowohl ein Kind, als auch mindestens ein Erwachsener und die Großeltern zusammenwohnen.

Die relativen Häufigkeiten der Altersklassen in Abbildung A.16 im Anhang zeigen deutlich, dass Haushalte mit Kindern im Alter bis 18 Jahre eher einen Pkw besitzen. In knapp 20 % der Haushalte mit Pkw gibt es ein Kind im Alter bis 15 Jahren, während in Haushalten ohne Pkw nur in gut 12 % mindestens ein Kind in dieser Altersklasse lebt. In der Altersklasse von 18 bis 30 Jahren überwiegt der Anteil an Haushalten ohne Pkw, was auf eine große Zahl an jungen Singlehaushalten mit Studierenden oder jungen Arbeitnehmern zurückzuführen sein könnte, die eher auf ein eigenes Auto verzichten. In der Altersgruppe zwischen 30 und 65 Jahre, die in beiden Gruppen den größten Anteil aller Altersgruppen einnimmt, besitzen mit knapp 49 % mehr Haushalte einen Pkw als keinen Pkw (42 %). In fast einem Viertel der Haushalte ohne Pkw sind Personen über 65 Jahre vorhanden, was die Haushalte mit Pkw mit über 15 % überwiegt.

Folglich lässt sich feststellen, dass in Haushalten mit Kindern und bei Erwachsenen ab 30 Jahren bevorzugt ein Pkw vorhanden ist. Haushalte mit jungen Erwachsenen und Rentnern sind eher den Alternativen zugewandt und besitzen daher kein Auto. Daraus folgt, dass eine Trennung anhand dieser Dummy-Variablen zu erwarten ist, weshalb diese in das Modell mit eingehen.

5.4.5 Erwerbstätigkeit

Grundsätzlich lässt sich die Erwerbstätigkeit als binäre Variable ausdrücken. In dieser Arbeit wird neben den Ausprägungen ‚ja' und ‚nein' zusätzlich das Attribut ‚in Ausbildung' betrachtet, um diese Gruppe auch berücksichtigen zu können. Es wird angenommen, dass sobald es in einem Haushalt eine erwerbstätige Person gibt, dies für den ganzen Haushalt gilt. Dabei wird vermutet, dass Haushalte mit erwerbstätigen Personen eher ein Auto besitzen, als Haushalte ohne erwerbstätige Person oder mit einer Person in Ausbildung.

Diese Vermutung bestätigen die relativen Häufigkeiten in Abbildung A.17 im Anhang. Bei den Haushalten mit Pkw sind genau 70 % erwerbstätig, während bei den Haushalten ohne Pkw nur 46 % einer Erwerbstätigkeit nachgehen. Bei den nicht erwerbstätigen Haushalten, zu denen unter anderem Seniorenhaushalte gehören, überwiegt der Anteil an Haushalten ohne Pkw (0,39) gegenüber den Haushalten mit Pkw (0,24). Ähnlich verhält es sich bei den Haushalten mit Personen, die sich in einer Ausbildung befinden. Insgesamt nimmt in beiden Gruppen die Ausprägung ‚erwerbstätig' den höchsten Anteil ein.

5.5 Nominale Variablen

Neben den bereits beschriebenen metrischen und kategorialen Variablen schließt sich eine Beurteilung der nominalen Variablen an, die die Ausprägungen null oder eins annehmen. Im folgenden wird kurz auf die Bedeutung eingegangen und anschließend mithilfe des korrigierten Kontingenzkoeffizienten (Kapitel 3.1.2) beurteilt, ob eine Aufnahme in das Modell sinnvoll ist.

Dienstwagen

Die Variable drückt aus, ob im Haushalt ein Dienstwagen vorhanden ist (ja = 1, nein = 0). Bei einem Dienstwagen ist von einer anderen Nutzungsweise auszugehen als bei privaten Pkw. Bei Dienstwagen besteht durchaus ein betriebliches Interesse an der Nutzung eines Pkws, wodurch der Haushalt selbst nicht die Entscheidung fällen muss, ob er sich privat ein Auto anschafft. Diese Variable führt zwangsläufig zum Pkw-Besitz bei Haushalten, da dieser durch die Variable impliziert wird. Darüber hinaus ist der Besitz eines zusätzlichen eigenen Pkw möglich. Dies ist in dieser Arbeit aber ohne Bedeutung. Ein Haushalt mit Dienstwagen gehört folglich zur Gruppe mit Pkw.

Verfügbarkeit ÖV-Haltestelle

Wie bereits in Kapitel 5.2.1 dargestellt, verfügen nicht alle Haushalte über einen Zugang zu einer Haltestelle. Diese wurden bei den Streu- und Lagemaßen mit einer Entfernung von 60 Minuten bewertet. Um eine weitere Unterscheidung zwischen den Haushalten, die über einen ÖV-Anschluss verfügen, zu gewährleisten, wird diese binäre Variable eingeführt. Diese nimmt den Wert eins an, wenn dem Haushalt eine ÖV-Haltestelle zur Verfügung steht, sonst null. Dabei ist zu erwarten, dass Haushalte ohne ÖV-Haltestelle wahrscheinlich einen Pkw besitzen.

Zugang zu schienengebundenen Verkehrsmitteln

Diese Variable wurde unter dem Aspekt generiert, den Schienenbonus zu berücksichtigen. Den

Wert eins nimmt die Variable an, wenn die Haltestelle mit der kürzesten Entfernung von einem Schienenverkehrsmittel des Nahverkehrs (Straßenbahn, U-Bahn) angefahren wird, sonst null. Zudem werden Haushalte berücksichtigt, die mit einer zusätzlichen Entfernung von drei Minuten gegenüber der nächsten (Bus-)Haltestelle ein Schienenverkehrsmittel erreichen. Dabei ist zu erwarten, dass aufgrund des meist dichten Angebotes dieser Verkehrsmittel die Nutzung von ÖV-Verkehrsmittel wahrscheinlicher ist.

Einschränkung

Eine wichtige Variable, die eine Präferenz gegenüber dem öffentlichen Verkehr bzw. Alternativen zum Pkw vermuten lässt, ist die Variable der körperlichen Einschränkung. Diese nimmt den Wert eins an, wenn eine körperliche Einschränkung vorliegt und den Wert null, wenn keine Einschränkung gegeben ist oder keine Angaben dazu gemacht wurden. Eine körperliche oder geistige Einschränkung lässt vermuten, dass diese Haushalte eine Präferenz gegenüber den Alternativen zum Auto haben, da diese oftmals kostenfrei genutzt werden können und die Pkw-Nutzung nur eingeschränkt möglich ist.

Besitz Dauerkarte

Der Besitz einer Dauerkarte (Wochen-, Monats- oder Jahreskarte) bei volljährigen Haushaltsmitgliedern lässt eine Bevorzugung des ÖV erwarten, da außer für die Fahrkarte keine weiteren Kosten für die Mobilität entstehen. Diese Präferenz und damit erwartbare Trennwirkung soll ebenfalls im Modell präsentiert werden. Liegt im Haushalt eine Dauerkarte vor, wird der Variable der Wert eins zugeordnet, sonst der Wert null.

Führerscheinbesitz

Ein Führerschein kann ein eindeutiges Indiz für die Zuordnung zu einer der beiden Gruppen sein. Besitzt in einem Haushalt niemand einen Führerschein, so ist die Wahrscheinlichkeit mindestens einen Pkw zu besitzen sehr gering. In diesem Falle nimmt die Variable den Wert null an. Der Besitz eines Führerscheines spricht eher für den Pkw-Besitz, wobei dies nicht zwingend korreliert. In dieser Arbeit wird zwischen dem Besitz eines Pkw-Führerscheins und eines Führerscheins für motorisierte Zweiräder unterschieden und jeweils mit einer eigenen Variablen modelliert.

Nutzung Bikesharing oder Carsharing

Neben den Alternativen zum Auto, wie der öffentliche Nahverkehr oder Zweiräder, besteht auch die Möglichkeit ein eigenes Autos durch Nutzung von Carsharing oder Bikesharing zu substituieren. Dabei ist davon auszugehen, dass Haushalte, die von Carsharing Gebrauch machen, eher keinen eigenen Pkw besitzen. Dieser Zusammenhang lässt sich in schwächerer Weise auch bei der Nutzung von Bikesharing als Ergänzung zum ÖV oder als Alternative zum Pkw vermuten. Bei der Nutzung dieser Sharingmöglichkeiten nimmt die Variable den Wert eins an, sonst null.

Zwischenstation in der Relation Arbeiten-Wohnen/Wohnen-Arbeiten, Beginn der Fahrt in der Schwachlastzeit

Diese Variablen wurden bereits im Abschnitt unter 4.2.2 erläutert. Liegt diese Eigenschaft in

den Haushalten vor, so nehmen die Variablen den Wert eins an, ansonsten null. In diesen Fällen ist eine Tendenz für die Kategorie des Pkw-Besitzes zu erwarten, da größere zeitliche und räumliche Flexibilität entscheidende Faktoren für die Rechtfertigung des Pkw-Besitzes sind. Eine Beurteilung anhand des Kontingenzkoeffizienten erscheint folglich sinnvoll.

Begleitung von Personen

Bei der Untersuchung der metrischen Variablen wurde in Abschnitt 5.2.2 die Anzahl der begleiteten Personen untersucht. Zusätzlich zur Anzahl wird über eine binäre Variable eine vorliegende Begleitung modelliert und im Modell berücksichtigt. Sofern eine Begleitung auf einem Weg vorliegt, nimmt diese den Wert eins an, sonst null. Es wird dabei angenommen, dass bei regelmäßiger Begleitung die Nutzung und damit der Besitz eines Pkws wahrscheinlicher ist.

5.6 Beurteilung der Variablen anhand des korrigierten Kontingenzkoeffizienten nach Pearson

Neben der sachlogischen und theoretischen Variablenauswahl erfolgt eine Beurteilung über die Aufnahme in das Modell über den korrigierten Kontingenzkoffizient nach Pearson, dessen Methodik im Abschnitt 3.1.2 dargestellt ist. In der folgenden Tabelle sind die Werte ersichtlich, die sich aus der Berechnung des Zusammenhangs zwischen den einzelnen nominalen Variablen und der abhängigen Variablen des Pkw-Besitzes ergeben.

Binäre Variable	K_*	**Binäre Variable**	K_*
Dienstwagen	0,242	Verfügbarkeit ÖV	0,004
Schienengebundene Verkehrsmittel	0,083	**Einschränkung**	0,171
Führerschein Pkw	0,594	**Führerschein Zweirad**	0,272
Besitz Dauerkarte	0,343	Nutzung Bikeharing	0,005
Nutzung Carsharing	0,172	Nachtzeit	0,057
Zwischenstation Wohnen-Arbeiten/Arbeiten-Wohnen	0,019 / 0,021	**Begleitung**	0,188

Tabelle 5.1: Korrigierter Kontingenzkoeffizient für nominalskalierte Variablen (e.D.)

Zusätzlich zu den nominal skalierten Variablen ist eine Untersuchung der kategorialen und ordinal skalierten Variablen mithilfe des korrigierten Kontingenzkoeffizienten möglich. Die Tabelle mit den zugehörigen Ausprägungen und Werten ist im Anhang in der Tabelle A.4 ersichtlich.

Im Modell in Kapitel 6 sollen nur Variablen berücksichtigt werden, deren korrigierter Kontingenzkoeffizent größer als 0,10 ist, um eine Wirkung und damit eine bessere Anpassung des Modells an die Daten zu gewährleisten. Diese Vorgehensweise wird auch unter dem Gebot der Sparsamkeit erfüllt, welches bei der logistischen Regression zu beachten ist (vgl. Backhaus et al. 2016, S. 333). Demnach liefert ein einfacheres Modell bessere Ergebnisse mit Daten außerhalb der Stichprobe. Die Variablen, die dieses Kriterium erfüllen, sind in den Tabellen 5.1 und A.4 im Anhang fett hervorgehoben.

6 Binäres Logit-Modell

Durch die in Kapitel 3.2 vorgestellte Methodik nach Backhaus et al. (2016, S. 284–334) lässt sich das Modell der binären logistischen Regression schätzen. Die im Modell aufgrund der deskriptiven Analyse berücksichtigten Variablen sind in Abbildung A.5 im Anhang aufgeführt. Das Modell basiert auf der abhängigen Variable mit den Ausprägungen ‚0 Pkw' und ‚mindestens 1 Pkw' auf Haushaltsebene. Im Folgenden wird ein Ausgangsmodell dargestellt und mithilfe von Optimierungsfunktionen verbessert. Anschließend wird das Modell unter Anwendung von verschiedenen Gütemaßen bewertet. Darüber hinaus wird eine Residuenanalyse durchgeführt. Abschließend erfolgt die Interpretation der sich ergebenden Regressionskoeffizienten und der zugehörigen Effektkoeffizienten. Für die Schätzung des Modells wurde eine Lernstichprobe generiert, der 70 % der Daten über einen Zufallsgenerator zugewiesen wurden. Einige Gütemaße werden mit Testdaten bestimmt, die die restlichen 30 % der Daten enthält.

6.1 Schätzung der Regressionskoeffizienten

Nach der Aufbereitung der metrischen und nominalen Variablen muss eine Bewertung der fehlenden Werte erfolgen. Für eine Schätzung der Parameter dürfen die zugrunde liegenden Datensätze keine fehlenden Werte enthalten. Dabei können entweder die betreffenden Datensätze aus dem Modell eliminiert oder, mithilfe von statistischen Methoden, die fehlenden Daten aufgefüllt werden. In dieser Arbeit wurde die zweite Variante angewendet und die fehlenden Werte mit dem Mittelwert der übrigen vorhandenen Werte ersetzt, um einen vollständigen Datensatz für die Modellierung zu generieren. Die Schätzung des Modells erfolgt mithilfe der Statistiksoftware R 3.5.1. Ausgehend vom Ausgangsmodell wird durch Eliminierung einzelner Variablen das Modell verbessert. Die Beurteilung der Güte erfolgt über das AIC, das in Kapitel 3.2.4 methodisch eingeführt wurde. Zudem wird der LL-Wert für die Beurteilung verwendet. Gemäß Tabelle 6.1

Variablen	**LL-Wert**	**AIC**
Ausgangsmodell	-5.646	11.346
Eliminierung Variable Alterklasse 18-29 Jahre	-5.645	11.344
Eliminierung Variable Altersklasse 30-65 Jahre	-5.647	11.344
Eliminierung Variable Realschulabschluss	-5.647	11.342
Eliminierung Variable Hochschulabschluss	-5.647	11.340

Tabelle 6.1: Optimierungsmöglichkeit des Logit-Modells (e.D.)

ergibt sich ein Modell mit 22 Variablen und einem AIC-Wert von 11.340. Dieses Ergebnis kann bei dieser großen Anzahl an Daten in der Stichprobe als gut eingeordnet werden.

	Variable	Parameter	Odds-Ratio	Relatives Risiko [1]	p-Wert (Wald-Test)
	Interzept (alternativenspezifische Konstante)	-3,755	0,023		0,000
Kardinale Daten	Wegeanzahl	-0,022	0,978		0,005
	Wegelänge	0,012	1,012		0,000
	Anzahl Personen	0,813	2,255		0,000
	Entfernung ÖV-Haltestelle	0,015	1,015		0,005
	Einkommen	0,001	1,001		0,000
	Durchschnittsalter	0,006	1,006		0,004
	Anzahl motorisierter Zweiräder	0,420	1,521		0,000
Nominale Daten	Einschränkung	-0,524	0,592	0,982	0,000
	Verfügbarkeit Dienstwagen	16,292	» 1	1,142	0,907
	Pkw-Führerschein	2,761	15,811	1,334	0,000
	Sonstiger Führerschein	0,463	1,589	1,012	0,000
	Besitz Dauerkarte	-1,262	0,283	0,964	0,000
	Nutzung Carsharing	-1,904	0,149	0,873	0,000
	Begleitung anderer Personen	0,161	1,174	1,005	0,024
Kategoriale Daten	Hauptschulabschluss	-0,310	0,733	0,990	0,001
	Abitur	-0,483	0,617	0,987	0,000
	Meisterabschluss	0,442	1,555	1,012	0,000
	Lehre	0,316	1,372	1,008	0,000
	Altersklasse 0-14 Jahre	-0,764	0,466	0,974	0,000
	Altersklasse 15-17 Jahre	-0,621	0,537	0,977	0,000
	Geschlecht männlich	-0,178	0,837	0,995	0,020
	Geschlecht weiblich	-0,298	0,742	0,991	0,000

[1] Relatives Risiko nur für nominale und kategoriale Variablen berechnet.

Tabelle 6.2: Ergebnisse des Logit-Modells (e.D.)

Die Schätzung dieser Parameter der Regressionsfunktion erfolgt unter Anwendung der Maximum-Likelihood-Funktion (Formel 3.16). Gleichzeitig stellen die in der Tabelle 6.2 genannten Parameter die Koeffizienten der Nutzenfunktion (Formel 3.9) dar. Mithilfe dieser Nutzenfunktion lässt sich durch Einsetzen der Daten die Wahrscheinlichkeit über die Formel 3.10 berechnen. Anhand dieses Wahrscheinlichkeitswertes lassen sich die Haushalte zu einer der beiden Gruppen einordnen. Die Klassifizierung der Elemente erfolgt unter Berücksichtigung eines Trennwertes. Dieser entspricht dem Anteil der Gruppe mit Pkw-Besitz in den Testdaten an der Gesamtzahl der Beobachtungen.

6.2 Prüfung des Gesamtmodells

Nach der Schätzung des binären logistischen Modells wird die Güte dieses Modells, wie in Kapitel 3.2.4 beschrieben, bewertet. Die einzelnen Prüfungsmethoden werden im Folgenden dargestellt. Eine Übersicht aller Gütemaße befindet sich im Anhang in Tabelle A.6.

6.2.1 Informationskriterien und Log-Likelihood-Wert

Diese Gütekriterien sind im Kapitel 3.2.4 näher beschreiben und dienen in erster Linie dem Vergleich zwischen verschiedenen Modellen. Im hier dargestellten Endmodell beträgt der AIC = 11.340, der BIC = 11.518 und der LL-Wert = -5.647. Aussagen über die Güte des Modells anhand der einzelnen Werte lassen sich nicht treffen. Es gilt auch hier, dass je niedriger der Wert bei den Informationskriterien bzw. höher beim LL-Wert, desto besser ist ein Modell zu beurteilen.

6.2.2 Likelihood-Ratio-Test

Der wichtigste Test zur Prüfung der Modellgüte ist der unter Kapitel 3.2.4 beschriebene Likelihood-Ratio-Test.

Im der Arbeit zugrunde liegenden Modell beträgt der Wert der LLR 6.154,4 und ist unter der Nullhypothese H_0: $\beta_1 ... \beta_{22} = 0$ angenähert χ^2-verteilt mit 22 Freiheitsgraden. Vergleicht man diesen Wert mit dem tabellierten χ^2-Wert für $\alpha = 0{,}05$ und 22 Freiheitsgraden mit LLR = $\chi^2_{emp} = 6.154{,}4 > 33{,}20$, so ist die Nullhypothese abzulehnen. Das Modell kann als statistisch signifikant beurteilt werden. Dem p-Wert als empirisches Signifkanzniveau kann der Wert 0,000 zugewiesen werden, sodass das Modell als höchst signifikant beurteilt werden kann.

6.2.3 Pseudo-R^2-Statistiken

Die Pseudo-R^2-Statistiken wurden in Kapitel 3.2.4 dargestellt und ergeben im geschätzten Modell folgende Werte: Der Wert für McFadden's R^2 beträgt 0,353, was unter Zuhilfenahme der Interpretationstabelle 3.2 für eine akzeptable Modellanpassung steht. Ebenfalls eine annehmbare Anpassung bescheinigt der Wert für das Cox & Snell-R^2 mit 0,307. Ein Wert von 0,475 für das Nagelkerke's R^2 attestiert dem Modell eine gute Anpassung, die nahe einer sehr guten Anpassung liegt. Folglich lässt sich das Modell anhand der Pseudo-R^2-Statistiken mit einer akzeptablen bis guten Güte bewerten, sodass erwartet wird, dass das Modell annehmbare Ergebnisse liefern wird.

6.2.4 Klassifizierung neuer Elemente

Ein weiteres Gütemaß für strukuren-prüfende Verfahren ist die Klassifizierungstabelle (Kapitel 3.2.4) und die sich daraus ergebende Trefferquote. Für die Einordnung der Elemente wird ein Trennwert benötigt. Im Testdatensatz gibt es 5.615 Haushalte mit Pkw-Besitz von 7.185 Haushalten im Datensatz. Dies entspricht einem Anteil von 0,781. Dieser Wert wird als Trennwert für die Einordnung verwendet. Mit dieser Trefferquote ergibt sich ein PCC (Formel 3.21) von 0,658. Für 0,781 als Trennwert ergibt sich folgende Klassifizierungstabelle:

		Prognostizierte Gruppenzugehörigkeit	
Tatsächliche Gruppenzugehörigkeit	Gruppe	Kein Pkw	1 oder mehr Pkw
	Kein Pkw	1.240 (79,0 %) Spezifität	330 (21,0 %)
	1 oder mehr Pkw	1.183 (21,1 %)	4.432 (78,9 %) Sensitivität

Tabelle 6.3: Klassifizierungsmatrix (e.D.)

Bei insgesamt 5.672 von 7.185 richtig zugeordneten Elementen errechnet sich eine globale Trefferquote von 0,789, die damit über dem Wert von PCC mit 0,658 liegt und damit gemäß Tabelle 3.2 eine bessere Zuordnung als bei zufälliger Einordnung ergibt. Damit ist das Modell positiv zu bewerten.

6.2.5 ROC-Kurve

Die Klassifizierungstabelle in Kapitel 6.2.4 ermittelt die Trefferquote für einen konkreten Trennwert. Dieser kann beliebig gewählt werden. Um die verschiedenen Trennwerte zu berücksichtigen und um eine allgemeinere Aussage treffen zu können, lässt sich die ROC-Kurve berechnen. Diese enthält alle möglichen Trennwerte zwischen null und eins, wonach beurteilt werden kann, bei welchem Trennwert das beste Ergebnis erzielt werden kann. Die ROC-Kurve ist in Abbildung 6.1 dargestellt.

Jeder Punkt auf der Linie stellt einen Trennwert dar. Die Klassifizierungstabelle mit dem Trennwert von 0,781 wird in Abbildung 6.1 durch den Schnittpunkt der blauen Linien repräsentiert. Zusätzlich lässt sich anhand dieser Abbildung die beste erreichbare Trefferquote für den Pkw-Besitz ablesen, die in diesem Modell bei etwa 0,84 (Sensitivität) liegt und vom Schnittpunkt der roten Linien repräsentiert wird. Mit dem Ablesen der Spezifität von etwa 0,76 kann eine Trefferquote von etwa 0,8 erwartet werden, die deutlich über dem Wert von PCC mit 0,658 liegt.

Neben der Trefferquote lässt sich der AUC-Wert berechnen, der im vorliegenden Modell bei 0,873 liegt. Dieser Wert ist nach Backhaus et al. (2016, S. 302) exzellent.

Betrachtet man die einzelnen Ergebnisse über alle Güteprüfungen hinweg, lässt sich das Modell als ‚gut' einordnen, woraus annehmbare Ergebnisse zu erwarten sind.

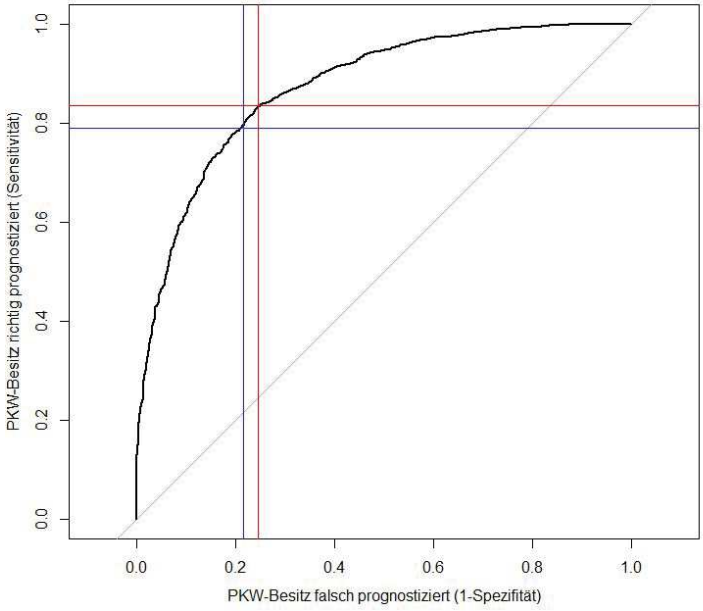

Abbildung 6.1: ROC-Kurve (e.D.)

6.3 Prüfung der Merkmalsvariablen

Zur Prüfung der Merkmalsvariablen bietet sich der Wald-Test und der LR-Test an (Kapitel 3.2.5). Die Ergebnisse für den Wald-Test sind in Tabelle 6.2 enthalten. Dabei weisen alle Parameter mit Ausnahme der Variable ‚Verfügbarkeit Dienstwagen' (0,907) eine hohe Signifikanz mit einer Irrtumswahrscheinlichkeit von $\alpha = 0{,}05$ auf.

Für diese Variablen wurde zusätzlich ein Likelihood-Quotienten-Test durchgeführt, um die Variable auf das Hauck-Donner-Phänomen zu prüfen, das bei Parametern mit sehr großem β_j auftritt (vgl. Wollschläger 2017, S. 317). Dabei sind die Streuungen deutlich zu groß, sodass die Parameter keine Signifikanz aufweisen. Die mangelnde Signifikanz entspricht aber nicht der Realität. Bei der Variable ‚Dienstwagen' ergibt der Likelihood-Quotienten-Test einen p-Wert von 0,000, wodurch diese Variable signifikant ist und das Hauck-Donner-Phänomen hier auftritt.

Insgesamt sind fast alle Parameter hoch signifikant bei einem Signifikanzniveau von 95 %. Dies lässt eine positive Wirkung auf die Güte des Modells erwarten.

6.4 Residuen-Analyse

In Kapitel 3.2.6 sind die methodischen Grundlagen zur Residuen-Analyse beschrieben, die in diesem Abschnitt am Lerndatensatz dargestellt werden. Eine Darstellung der Residuen findet sich in einem Streudiagramm in Abbildung 6.2.

In diesem Zusammenhang lassen sich alle Punkte, die absolut größer als die standardisierte Standardabweichung von zwei sind, als Ausreißer einordnen. Insgesamt können in der Lernstichprobe 740 Haushalte identifiziert werden, die sich nach dieser Definition als Ausreißer bezeichnen lassen. Das entspricht einem Anteil von 4,4 %.

Bei stichprobenartigen Überprüfungen der als Ausreißer bezeichneten Haushalte fällt auf, dass

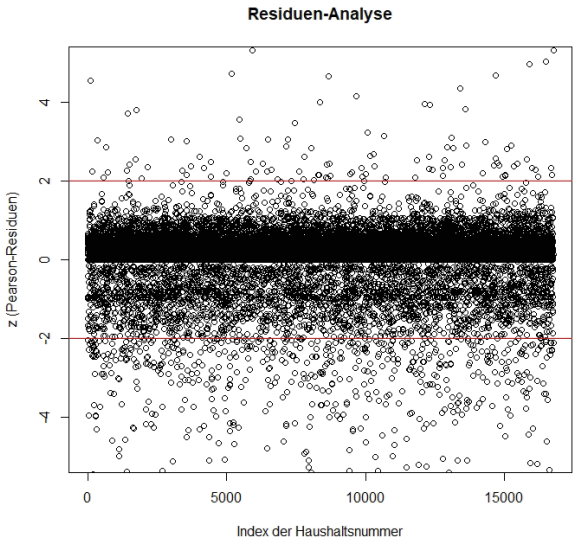

Abbildung 6.2: Standardisierte Residuen (e.D.)

diese Abweichungen als plausibel eingestuft werden können. Wird beispielsweise der Haushalt mit der Identifikationsnummer 414092 untersucht, so wohnen in diesem Haushalt sechs Personen und besitzen trotzdem keinen Pkw, aber zwei motorisierte Zweiräder. Dies erscheint plausibel, sodass hier nicht von Fehlern in der Erhebung ausgegangen werden muss. Bei der Untersuchung weiterer Ausreißer ergaben sich ähnliche Ergebnisse. Unter der Annahme, dass die logistische Regression relativ robust gegenüber Ausreißern ist und die identifizierten Ausreißer als plausibel zu beurteilen sind, empfiehlt es sich, diese nicht aus dem Modell zu eliminieren (vgl. Backhaus et al. 2016, S. 321). Bei dem großen Stichprobenumfang ist zudem bei Eliminierung der Parameter keine starke Änderung der Koeffizienten zu erwarten, da sich die Extreme in verschiedenen Variablen widerspiegeln.

6.5 Interpretation und Diskussion der Regressionskoeffizienten

Ziel dieser Arbeit ist es unter anderem, die Eigenschaften eines Haushaltes zu analysieren, die den Pkw-Besitz eines Haushaltes beschreiben. Zudem soll beurteilt werden, wie bestimmte Ereignisse die Pkw-Ausstattung beeinflussen können. Diese Beschreibung kann mithilfe der durch das binäre Logit-Modell berechneten Ergebnisse geschehen.

6.5.1 Metrische Variablen

Die Interpretation der kardinal skalierten Merkmalsvariablen erfolgt mithilfe des OR (Effekt-Koeffizient, siehe Formel 3.33). Die Werte des OR sind in Tabelle 6.2 ersichtlich. Einen großen Einfluss auf die Vorhersagbarkeit von Pkw-Besitz hat die Anzahl der Personen pro Haushalt. So steigt die Chance einen Pkw zu besitzen mit einer zusätzlichen Person um 2,257, was den höchsten Wert aller kardinalen Variablen entspricht. Folglich hat die Personenanzahl pro Haushalt den größten Einfluss auf den Pkw-Besitz.

Eine zunehmende Chance auf den Pkw-Besitz ist zudem bei der Variable ‚Einkommen' gegeben. Hier ist zu beachten, dass der OR-Wert nur eine geringe Zunahme der Chance um 1,001 mit jedem zusätzlichen Euro an Einkommen widerspiegelt. Da in dieser Arbeit die Klassenmitte betrachtet wird, erhöht sich der Sprung von einer Gruppe zur anderen um viele hundert Euro und wirkt sich damit auf die Aussage über den Pkw-Besitz aus. Die Chance ist demnach wesentlich höher bei steigendem Einkommen. Das ist unter dem Aspekt nachvollziehbar, dass die Anschaffung eines kapitalintensiven Autos mit steigendem Einkommen einfacher zu bewältigen ist. Diese Vermutung wurde bereits durch die deskriptive Analyse bestätigt.

Ebenso einen positiven Einfluss auf die Chance eines Haushalts, ein Auto zu besitzen, hat das Durchschnittsalter. Mit jedem zusätzlichen Jahr erhöht sich die Chance um 1,006. Dies widerspricht den Ergebnissen aus der deskriptiven Analyse insoweit, dass im Rentenalter der Anteil der Haushalte ohne Pkw überwiegt. Eine Begründung dafür könnte sein, dass in der Stichprobe der Anteil von Haushalten mit einem Durchschnittsalter über 65 Jahren im Gegensatz zur Altersklasse von 30 bis 65 Jahren relativ gering ist.

Weiterhin lässt sich eine erhöhte Chance für den Pkw-Besitz bei der Variable ‚Wegelänge' ablesen. Mit einer Zunahme der täglichen Wegelänge um einen Kilometer pro Tag, erhöht sich die Chance auf den Pkw-Besitz um 1,012. Dies bestätigt die Annahme, dass Haushalte mit längeren Wegen eher über einen Pkw verfügen.

Ebenso verhält es sich bei der Entfernung zur nächsten ÖV-Haltestelle. Je weiter die zeitliche Entfernung zur nächsten Haltestelle ist, desto eher ist der Haushalt im Besitz eines Pkws, was einer realitätsnahen Annahme entspricht.

Haushalte mit Pkw-Besitz lassen sich ebenfalls über den Besitz von motorisierten Zweirädern charakterisieren. So erhöht sich die Chance für den Pkw-Besitz mit jedem zusätzlichen motorisierten Zweirad um 1,523. Demnach kann vermutet werden, dass Haushalte mit einem Pkw eine gewisse Affinität gegenüber motorisierten Verkehrsmitteln haben und daher ein Motorrad eher als Zusatzausstattung statt als Alternative zum Auto angesehen wird. Dieser recht hohe Effekt-Koeffizient bestätigt zudem den Eindruck aus der deskriptiven Statistik.

Die Ergebnisse des OR erlauben zudem eine Einordnung der metrischen Variablen, die Haushalte ohne Pkw charakterisieren. So ist die Chance keinen Pkw zu besitzen mit jedem zusätzlichen Weg um 0,978 größer. Dieser schwache Effekt überrascht, da nach der deskriptiven Analyse ein differenziertes Ergebnis zugunsten des Pkw-Besitzes zu erwarten war. Hier zeigt sich, dass die Ergebnisse zwischen der Modellschätzung und der deskriptiven Beschreibung auch variieren können. Eine mögliche Begründung für dieses Ergebnis könnte sein, dass Pkw-lose Haushalte nicht ausreichend Erledigungen auf einem Weg bündeln können, was an umständlicher Wegeführung oder mangelnder Transportkapazität liegen könnte.

6.5.2 Nominale Variablen

Für die Interpretation der nominalen und kategorialen Variablen wird neben dem OR auf das RR (Kapitel 3.2.7) zurückgegriffen. Bei metrischen Variablen hängt die Größe des OR von der Maßeinheit ab (vgl. Backhaus et al. 2016, S. 313). Da bei binären Variablen selten eine Einheit gegeben ist, wird dieses Problem mit dem RR umgangen. Die Ergebnisse für das RR sind in der Tabelle 6.2 enthalten. Für die Berechnung des RR wird bei den übrigen, nicht betrachteten

Variablen der Mittelwert der Ergebnisse aus der Lernstichprobe verwendet.

Bei den binär kodierten Variablen hat ein verfügbarer Dienstwagen die größten Chancen einen Pkw-Besitz zu prognostizieren. So sind das OR als auch das RR weit über eins. Dies ergibt sich zwangsläufig, da ein Haushalt mit Dienstwagen stets über einen Pkw verfügt und damit eindeutig der Gruppe mit Pkw-Besitz zuzuordnen ist.

Die Chancen für den Pkw-Besitz erhöht ebenfalls die Variable ‚Begleitung von Personen'. Haushalte, in denen Personen begleitet werden, besitzen 1,005-mal wahrscheinlicher einen Pkw als Haushalte, in denen keine Begleitung stattgefunden hat.

Das gleiche Bild lässt sich bei Haushalten mit Führerscheinbesitz ablesen. Mit Führerschein liegt die Chance für den Pkw-Besitz wesentlich höher als bei Haushalten ohne Führerschein. Diese Wirkung ist unabhängig davon, ob ein Pkw-Führerschein oder ein Führerschein für ein sonstiges Verkehrsmittel vorliegt, wobei der Effekt beim Pkw-Führerschein wesentlich größer ist.

Nominale Variablen, die die Chance für den Pkw-Besitz reduzieren bzw. für einen Haushalt ohne Auto stehen, sind der Besitz einer Dauerkarte, das Vorliegen einer Beschränkung und die Nutzung von Carsharing. Den stärksten Effekt davon weist die Variable ‚Nutzung von Carsharing' auf, gefolgt vom Dauerkartenbesitz und dem Vorliegen einer Beschränkung. So besitzt ein Haushalt, der Carsharing nutzt, 0,874-mal wahrscheinlicher mindestens einen Pkw als ein Haushalt, der Carsharing nicht nutzt. Folglich besitzen Haushalte, die Carsharing, nutzen eher keinen Pkw. Ein ähnlicher Effekt ergibt sich aus den beiden anderen Variablen. Dieses Ergebnis erscheint plausibel zu sein, da die Nutzung der Alternativen ein eigenes Auto überflüssig machen. Zudem führt das Vorliegen einer Mobilitätseinschränkung dazu, dass in der Regel eine Freifahrtberechtigung für den öffentlichen Verkehr vorliegt oder das eigene Fahren eines Pkws nur eingeschränkt möglich ist, wodurch Haushalte in denen eine Einschränkung vorliegt, in nachvollziehbarer Weise die Alternativen zum Pkw präferieren.

6.5.3 Kategoriale Variablen

Die Interpretation des RR erfolgt in Bezug auf die Referenzkategorie aus Tabelle A.4 im Anhang, die aufgrund der Multikollinearität in der Modellschätzung nicht enthalten ist. Bei der Untersuchung der kategorialen Variable des überwiegenden Geschlechts in einem Haushalt fällt auf, dass ein Haushalt mit überwiegend weiblichen Mitgliedern 0,991-mal wahrscheinlicher und Haushalte mit hauptsächlich männlichen Mitgliedern 0,995-mal wahrscheinlicher im Pkw-Besitz sind gegenüber Haushalten, die eine gleiche Anzahl beider Geschlechter haben. Zudem tendieren weiblich geprägte Haushalte eher dazu, keinen Pkw zu besitzen als männlich geprägte Haushalte.

Für die Interpretation der folgenden Merkmalsvariable muss eine Annahme getroffen werden. Durch die Optimierung des Ausgangsmodells werden einige Dummy-Variablen eliminiert, die für die Interpretation der kategorialen Merkmale von Bedeutung sind. Die Eliminierung beruht darauf, dass durch diese Dummy-Variablen keine eindeutige Wirkung ausgeht. Das wird durch die Berechnung der Konfidenzintervalle für ein Konfidenzniveau von 95 % deutlich. Alle Variablen, die im Ausgangsmodell gegenüber dem reduzierten Endmodell vorhanden sind, weisen diese Schwankungen beim OR auf. Eine Zusammenfassung von Variablen, wie beispielsweise kleinere Altersklassen, ergab den gleichen Effekt. Um auch Aussagen über diese Dummy-Variablen treffen zu können, werden die OR-Werte aus dem Ausgangsmodell verwendet, deren Robustheit

mit einem kleineren Modell geprüft wurde, um die Aussagekraft ansatzweise zu validieren. Die Ergebnisse dieser weitergehenden Betrachtung sind aus der Tabelle A.7 im Anhang ersichtlich.

Die Altersklassen lassen sich wie folgt interpretieren. Haushalte mit Kindern in der Altersgruppe zwischen 0 bis 14 und 15 bis 17 Jahren besitzen 0,973-mal bzw. 0,976-mal wahrscheinlicher einen Pkw als Haushalte, in denen Personen über 65 Jahren vorhanden sind. Haushalte mit Erwachsenen im Alter zwischen 30 bis 65 Jahren sind nahezu genauso wahrscheinlich im Pkw-Besitz wie Haushalte mit Rentnern. Dies ist insofern nachvollziehbar, dass Haushalte ab dem führerscheinfähigen Alter eher im Pkw-Besitz sind. Hier lässt sich auch erkennen, dass die Werte für das OR bei den Altersklassen zwischen 18 und 65 Jahren im Konfidenzintervall von 95 % keine eindeutige Einordnung erlauben und sowohl größer als auch kleiner als eins sein können. Daher wurden diese Werte aus dem endgültigen Modell eliminiert.

Die Dummy-Variable der Altersklassen 0 bis 14 und 15 bis 17 Jahre kann zusätzlich zur Interpretation verwendet werden, ob Kinder im Haushalt eher zum Pkw-Besitz führen oder nicht. In dieser Auswertung ergibt sich, dass Haushalte mit Kindern etwa 0,97-mal wahrscheinlicher im Pkw-Besitz sind als Haushalte ohne Kinder. Aus diesem überraschenden Ergebnis lässt sich die Annahme widerlegen, dass Haushalte mit Kindern eher im Pkw-Besitz sind. Dies könnte daran liegen, dass ein Haushalt im urbanen Gebiet meist kurze Wege für Erledigungen und ein gutes Angebot an Alternativen hat. Allerdings ist der Anteil an Haushalten mit Kindern in der Stichprobe nicht sehr groß, sodass es hierbei zu Verzerrungen kommen kann.

Bei der Variable ‚Höchster Schulabschluss' ergibt sich, dass eine eindeutige Aussage nur bei Haushalten getroffen werden kann, deren höchster Schulabschluss das Abitur ist. Diese Haushalte sind 0,987-mal wahrscheinlicher im Besitz eines Pkws als Haushalte ohne Schulabschluss. Das gleiche Ergebnis in schwächerer Form zeigt sich beim Vergleich zwischen Hauptschulabschluss und ohne Abschluss, allerdings schwankt auch hier das Konfidenzintervall stark. Keine Aussage lässt sich beim Realschulabschluss treffen. Bei diesem liegt das RR nahezu bei 1,000, weshalb auch diese Variable aus dem Modell eliminiert wurde. Grundsätzlich lässt sich aussagen, dass Haushalte mit einer höheren Schulbildung eher keinen Pkw besitzen, was der Vermutung aus der deskriptiven Analyse widerspricht. Dies lässt sich dadurch begründen, dass in diesen Haushalten, gegenüber Haushalten mit einer niedrigeren Schulbildung, der Pkw den Wert als Statussymbol verloren und sich ein größeres Umweltbewusstsein und Kostensensibilität entwickelt haben könnte. Zudem sind Personenkreise mit niedriger Bildung öfter an feste Arbeitszeiten gebunden, sodass bei Schichtarbeit zum Beispiel das Angebot an Alternativen geringer ist.

Ein leicht differenzierteres Bild liefert die Variable der ‚Berufsausbildung'. Hier besitzen Haushalte mit einer Lehre oder einem Meister als Abschluss eher ein Auto als Haushalte mit einem Hochschulabschluss. Diese Interpretation beruht auf Grundlage der Referenzkategorie von Haushalten ohne Abschluss. Allerdings tendieren in dieser Betrachtung Haushalte mit einem Hochschulabschluss eher in Richtung Pkw-Besitz gegenüber Haushalten ohne Berufsausbildung. Die Dummy-Variable des Hochschulabschlusses lässt aber keine eindeutige Interpretation zu, da auch deren Konfidenzintervall des OR zwischen den beiden Gruppen schwankt. Der Unterschied zwischen Haushalten mit Lehre und Haushalten mit Meisterabschluss ist marginal mit einer Tendenz zum Pkw-Besitz gegenüber Haushalten ohne Berufsausbildung. Der Effekt beim Hochschulabschluss ist dabei durchaus geringer, sodass hier eine größere Präferenz gegenüber

den Alternativen zu erwarten ist, was sich ebenfalls beim Schulabschluss herauskristallisiert hat und der gleichen Begründung folgt.

Neben dem OR konnten für die einzelnen Beobachtungen die Odds und die Logits berechnet werden. Eine Interpretation der einzelnen Elemente würde in dieser Arbeit zu weit führen. Für die Logits lässt sich aussagen, dass je höher der Wert für die Logits, desto höher ist auch die Wahrscheinlichkeit für den Pkw-Besitz. Aus den Logits der einzelnen Beobachtungen lässt sich der S-förmige Verlauf der Logit-Funktion ableiten, wie der Abbildung 6.3 zu entnehmen ist.

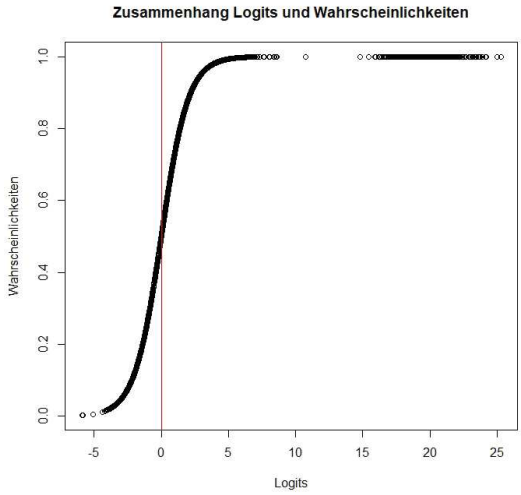

Abbildung 6.3: Zusammenhang zwischen den Logits und Wahrscheinlichkeiten der Beobachtungen (e.D.)

Alternativenspezifische Konstante

Die alternativenspezifische Konstante nimmt in diesem Modell den Wert -3,758 bzw. ein OR von 0,023 an und spiegelt damit eine globale Präferenz für die Kategorie der Haushalte ohne Pkw wider. Darin werden alle systematischen Einflussfaktoren berücksichtigt, die nicht über Variablen im Modell beschrieben werden.

6.5.4 Konfidenzintervalle

Zur Überprüfung der Konstanz der Wirkungen des OR wurde für alle einzelnen Variablen die Konfidenzintervalle für ein Signifikanzniveau von 95 % gebildet. Die Ergebnisse aus Tabelle A.8 im Anhang zeigen, dass die Effekte der Variablen in 95 % der Fälle konstant sind. Die Interpretationsergebnisse der Variablen sind als gut zu interpretieren.

7 Fazit

Ziel dieser Arbeit ist die Analyse und die Charakterisierung des Pkw-Besitzes in Haushalten. Anhand der Ergebnisse der deskriptiven Analyse und des binären logistischen Modells lassen sich Haushalte im Bezug des Pkw-Besitzes folgendermaßen beschreiben:

Je mehr Personen in einem Haushalt leben und je höherer deren Einkommen ist, desto eher ist ein Pkw im Haushalt vorhanden. Zudem ist ein Haushalt mit täglich langen Wegen eher im Besitz eines Pkws als ein Haushalt mit kürzeren Wegen. Auch ein motorisiertes Zweirad weist sehr wahrscheinlich auf den Besitz eines Autos im Haushalt hin. Den gleichen Effekt erzielt der Führerscheinbesitz, wobei es keinen Unterschied macht, ob ein Pkw-Führerschein oder ein sonstiger Führerschein vorliegt. Neben dem Führerscheinbesitz spielt auch das Durchschnittsalter eine Rolle beim Pkw-Besitz, indem ein höheres durchschnittliches Alter eher auf einen Pkw-Besitz schließen lässt. Außerdem befindet sich in diesen Haushalten häufig eine Person im Alter zwischen 30 und 65 Jahren. Zusätzlich ist auch die Entfernung zur nächsten Haltestelle in schwachem Maße bedeutend, denn je weiter ein Haushalt von der nächsten Haltestelle entfernt liegt, desto eher ist ein Auto verfügbar. Ein männlich geprägter Haushalt verfügt zudem eher über einen Pkw als ein weiblich geprägter Haushalt. Darüber hinaus ist die Wahrscheinlichkeit für den Pkw-Besitz in Haushalten ohne vorherrschendes Geschlecht größer als bei weiblichen oder männlichen Haushalten. Die Verfügbarkeit eines Dienstwagens weist schließlich deutlich auf einen Pkw-Besitz hin.

Die Wahrscheinlichkeit, dass ein Haushalt keinen Pkw besitzt, wird durch das Vorliegen der Eigenschaften ‚Besitz einer Dauerkarte', ‚Nutzung von Carsharing' und ‚Kurzer Weg zur nächsten Haltestelle' erhöht. Zudem besitzen Haushalte, in denen mindestens eine Person mit einer körperlichen Einschränkung lebt, tendenziell auch keinen Pkw. Je höher der Schulabschluss eines Haushaltes, desto wahrscheinlicher ist der Verzicht auf einen Pkw, wobei dieser Effekt als marginal zu beurteilen ist. Ähnlich verhält es sich bei der Berufsausbildung: Haushalte mit Lehre oder Meister als Berufsabschluss besitzen häufiger ein Auto. Darüber hinaus verfügen laut dem Modellergebnis Haushalte mit einer größeren Anzahl an Wegen eher über keinen Pkw. Das ist genauso überraschend wie die Eigenschaft, dass Haushalte mit Kindern bis 18 Jahren eine leichte Tendenz dazu aufweisen, keinen Pkw zu haben.

Ein eindeutiger Pkw-Haushalt besitzt die modellmäßige Wahrscheinlichkeit von nahezu eins. Aus der Nutzenfunktion gemäß Tabelle 6.2 ergibt sich ein Wert in Höhe von 25,338.

$$p(Y = 1) = \frac{1}{1 + e^{25{,}338}} \approx 1 \tag{7.1}$$

Im eindeutigen Pkw-Haushalt leben fünf Personen, darunter Kinder bis 14 Jahren, und es ist ein Dienstwagen verfügbar. Die Entfernung zur nächsten Haltestelle beträgt drei Minuten. Das monatliche Haushaltseinkommen liegt über 5.600 € und das Durchschnittsalter liegt bei

33,5 Jahren. Der höchste Schulabschluss in diesem Haushalt ist ein Realschulabschluss und als Berufsausbildung liegt ein Meister vor. Der Haushalt besitzt zudem einen Pkw-Führerschein und einen sonstigen Führerschein sowie keine Dauerkarte für den ÖV. Zudem erfolgt keine Begleitung und keine Nutzung von Carsharing. Insgesamt werden täglich 22 Wege mit einer Länge von 194 Kilometer zurückgelegt.

Aus der Nutzenfunktion ergibt sich für den klassischen Haushalt ohne Pkw, also mit der geringsten vorhergesagten Wahrscheinlichkeit, der Wert -5,826.

$$p(Y = 1) = \frac{1}{1 + e^{-5,826}} \approx 0 \tag{7.2}$$

Der eindeutige Haushalt ohne Pkw ist charakterisiert durch einen weiblichen Singlehaushalt, ein geringes mittleres Einkommen von 700 € und einer Entfernung von sieben Minuten zur nächsten Haltestelle. In diesem Haushalt gibt es kein motorisierter Verkehrsmittel. Das Durchschnittsalter liegt bei 41 Jahren und es liegt eine körperliche Einschränkung vor. Der höchste Schulabschluss ist ein Hauptschulabschluss auf den eine Lehre als Berufsabschluss folgte. Es gibt zudem im Haushalt keinen Führerschein. Außerdem verfügt der Haushalt über eine Dauerkarte, nutzt Carsharing und begleitet andere Personen außerhalb des Haushalts. Die tägliche Wegelänge beträgt 17 Kilometer, die in drei Wegen absolviert wird.

Dabei fällt auf, dass die Angaben der Person einen Widerspruch darstellt, da Carsharing genutzt wird, obwohl die Person keinen Führerschein hat. Dies könnte auf falsche Angaben zurückzuführen sein oder auf andere Konstellationen wie dem Führen des Pkw durch Bekannte des Haushaltes.

8 Diskussion und Literatur

Werden die Ergebnisse aus der Literatur (Kapitel 2) mit den Erkenntnissen aus dieser Arbeit verglichen, so lassen sich einige Rückschlüsse ziehen und Parallelen erkennen.

Viele Parallelen zu dieser Arbeit sind in Van Acker & Witlox (2010) enthalten. Beide Werke befassen sich mit dem Pkw-Besitz in Haushalten. Die Veröffentlichung von Van Acker & Witlox (2010) befasst sich, im Gegensatz zu dieser Ausarbeitung, nicht mit dem Pkw-Besitz als rein abhängige Variable, sondern betrachtet insbesondere die Bedeutung des Pkw-Besitzes als intervenierende Variable zur Vorhersage der Pkw-Nutzung. In diesem Modell dient beispielsweise die Wegelänge als exogene Variable, während Van Acker & Witlox (2010) die Wegelänge als endogene Variable unter Beachtung des Pkw-Besitzes prognostizieren. Zudem stehen in diesem Werk die Eigenschaften des Pkw-Besitzes im Vordergrund, während Van Acker & Witlox (2010) die Pkw-Nutzung als endogene Variable betrachten. Auch das für die Analyse verwendete Modell unterscheidet sich in beiden Untersuchungen. Während die Modellierung des Pkw-Besitzes in dieser Arbeit auf der binären logistischen Regression beruht, verwenden Van Acker & Witlox (2010) ein Strukturgleichungsmodell, in dem simultan verschiedene Regressionsgleichungen geschätzt werden, um die Effekte zwischen den Variablen abbilden zu können. Einen weiteren Unterschied bildet die Datengrundlage. Die Daten von Van Acker & Witlox (2010) liegen auf Personenebene vor, während diese Arbeit auf Haushaltsinformationen basiert. Dazu ist die abhängige Variable in dieser Ausarbeitung nur binär mit den Ausprägungen ‚0 Pkw' und ‚1 oder mehr Pkw' definiert, während die Studie differenzierter ist und zusätzlich die Ausprägung ‚1 Pkw' separat in die Betrachtung mit aufnimmt. Dadurch liegen die Ergebnisse bei Van Acker & Witlox (2010) wesentlich näher an den einzelnen Personen, als die Erkenntnisse dieser Arbeit. Letztlich lassen sich die Resultate beider Arbeiten unter diesem Aspekt nur eingeschränkt vergleichen. Werden die Ergebnisse beider Werke trotzdem gegenübergestellt und verglichen, fällt auf, dass die Entfernung zur nächsten Haltestelle, das Einkommen, die Haushaltsgröße und der Führerscheinbesitz in beiden Studien ähnliche Rückschlüsse ergeben. Ein leicht unterschiedliches Ergebnis ergibt sich beim Alter. In dieser Analyse spricht ein höheres Durchschnittsalter eher für den Besitz eines Pkws, während Van Acker & Witlox (2010) mit höherem Alter eher keinen Pkw-Besitz prognostizieren. Dies könnte durch die Multilevelstruktur bedingt sein, durch die gewisse Informationen verzerrt werden. Zudem sind die Variablen ‚Geschlecht', ‚Bildung' und ‚Kinder' im Modell von Van Acker & Witlox (2010) mangels Signifikanz aus dem Modell eliminiert worden, während diese Variablen in dieser Untersuchung eine hohe Signifikanz aufweisen. Ein Vergleich ist daher nicht möglich. Van Acker & Witlox (2010) betrachten zusätzlich die bauliche Umgebung der Personen in ihrer Analyse und ziehen Rückschlüsse daraus. Mangels vorliegender Regionaldaten kann eine derartige Untersuchung in dieser Arbeit nicht erfolgen. Insgesamt ähneln sich die Ergebnisse der beiden Studien, wobei eine Betrachtung der Pkw-Variable als intervenierende Variable keinen Mehrwert bringen würde.

Ein inhaltlicher Vergleich zwischen den Erkenntnissen von Bhat & Sen (2006) und Collin-Lange & Benediktsson (2011) mit dieser Arbeit ist aufgrund der unterschiedlichen Fragestellungen kaum möglich. Es lassen sich aber Parallelen erkennen und eine Ergänzung der Ergebnisse ableiten.

In der Studie von Bhat & Sen (2006) wird der Pkw-Typ beschrieben, wobei ein Pkw-Besitz vorausgesetzt wird. In dieser Arbeit tendieren größere Familien zum Pkw-Besitz, während Bhat & Sen (2006) Aufschluss darüber geben, dass diese eher SUVs und Minivans bevorzugen. Ein grundsätzlicher Vergleich beider Studien erscheint auch aus der Hinsicht schwierig, da deutsche bzw. amerikanische Haushalte als Basis dienen. Beide Länder haben einen unterschiedlichen Bezug zum Pkw, sodass Aussagen nicht allgemeingültig auf Haushalte beider Länder übertragbar sind. Zusätzlich ist in der Veröffentlichung von Bhat & Sen (2006) der Einfluss von Betriebskosten berücksichtigt, die in der vorliegenden Arbeit nicht betrachtet werden, aber durchaus einen Einfluss haben könnten.

Collin-Lange & Benediktsson (2011) untersuchen in ihrer Studie den Stellenwert des Pkws unter Jugendlichen in Island. Ein Vergleich zwischen Ergebnissen aus Island mit denen in dieser Studie ist schwer möglich, da unterschiedliche Gegebenheiten vorliegen. Gemäß der Studie von Collin-Lange & Benediktsson (2011) können isländische Jugendliche ihren Pkw-Führerschein bereits mit 16 Jahren erwerben, während dies in Deutschland ohne Berücksichtigung von Sonderformen erst ab 18 Jahren möglich ist. Außerdem gibt es in Deutschland, vorwiegend in Städten, ein dichtes Angebot an ÖV, das eine Alternative zum eigenen Pkw darstellt. Die Ergebnisse dieser Studie zeigen, dass Haushalte mit Kindern bis 18 Jahren und junge Erwachsene bis 30 Jahren dazu tendieren, keinen Pkw zu besitzen, wobei die Altersklasse von 18 bis 29 Jahren keine signifikanten Ergebnisse aufweist. Dieser Vergleich ist auch kritisch zu beurteilen, da hier die Haushaltsebene mit der Personenebene verglichen wird. Insgesamt lässt sich aber erkennen, dass nach der vorliegenden Arbeit der Pkw-Besitz in Städten keinen großen Stellenwert als Statussymbol unter jungen Erwachsenen hat, da die Wahrscheinlichkeit auf den Pkw-Besitz mit zunehmenden Alter erst größer wird.

Zusammengefasst lässt sich sagen, dass die Ergebnisse dieser Arbeit im Vergleich mit anderen Erkenntnissen aus der Literatur als plausibel und wahrscheinlich eingestuft werden können. Zudem können die Ergebnisse anderer Studien die Erkenntnisse dieser vorliegenden Untersuchung ergänzen und helfen diese einzuordnen.

9 Kritische Würdigung und Ausblick

Zusammenfassend lässt sich sagen, dass sich mithilfe des binären Logit-Modells Charakteristika der Haushalte für den Pkw-Besitz ableiten lassen. Die Ergebnisse der Analyse aus Kapitel 7 stellen nur einen kleinen Ausschnitt möglicher Charakteristika zur Beschreibung des Pkw-Besitzes in Haushalten dar. In der Praxis sind noch viele weitere Einflussfaktoren denkbar, die in dieser Betrachtung unberücksichtigt bleiben. So ist es schwierig soziale Wirkungen, wie der bereits in der Literaturübersicht behandelte Einfluss als Statussymbol, empirisch zu erfassen und damit in das Modell mit aufzunehmen. Eine nicht weniger bedeutende Rolle spielt die bauliche Umgebung der Haushalte, die mangels Informationen aus der Umfrage nicht beurteilt werden können. Für eine fundierte Analyse scheint die Betrachtung dieser Variablen von großer Bedeutung. Ebenso wichtig ist die Bedienungsqualität des öffentlichen Verkehrs. In dieser Arbeit wurde zwar die Zugangszeit zur nächsten Haltestelle betrachtet, aber nicht wie oft und in welcher Qualität die Bedienung dieser Haltestelle erfolgt. Ebenso zu bewerten sind beispielsweise die Verfügbarkeiten von guten und sicheren Radwegen sowie kurze Wege für Erledigungen des täglichen Bedarfs. Dieser Mangel an Informationen hat auch einen Einfluss auf die Interpretation der Ergebnisse, da für die Haushalte keine Standortinformationen verfügbar sind und damit die Bewertung der Parameter der Alternativen an Aussagekraft verliert. Eine Großstadt wie Berlin gewährleistet in der Regel ein anderes Angebot an Verkehrsinfrastruktur als kleinere Städte wie Dessau-Roßlau. Andererseits kann aufgrund des breiten Spektrums verschiedener Städte mit unterschiedlichen Strukturen ein allgemeineres Bild kreiert und dieses auch auf andere Regionen übertragen werden.

Aufgrund der in der Datengrundlage vorliegenden Multilevelstruktur kann davon ausgegangen werden, dass einige Informationen durch die Aggregation auf Haushaltsebene verloren gehen oder deren Aussagekraft reduziert wird. Die Aggregation führt dazu, dass beispielsweise Wegeinformationen von Personen gelöscht werden für die kein normaler Tag vorgelegen hat. Diese Eliminierung ist der Multilevelstruktur geschuldet und würde nicht auftreten, wenn die Befragung rein auf Haushaltsebene stattgefunden hätte. Gleichzeitig könnte damit erreicht werden, dass die Anzahl fehlender Werte reduziert wird, die sich unter anderem durch die Komprimierung ergeben. Für aussagekräftigere Analysen und für mehr Informationen empfiehlt es sich daher, eine Stichprobe zu verwenden, in der bereits alle Informationen auf Haushaltsebene vorliegen und damit eine Aggregation von Daten samt Informationsverlust überflüssig macht.

Durch die Multilevelstruktur der Stichprobe ist auch eine binäre Modellierung der abhängigen Variable begründet. In einer Vorversion des Modells war die Modellierung als kategoriale Variable mit den Ausprägungen ‚Kein Pkw', ‚1 Pkw' und ‚2 oder mehr Pkw' angedacht, um weitere Charakteristika generieren zu können. Nach ersten Schätzversuchen ist das Modell an der Multikollinearität der Daten selbst dann gescheitert, wenn nur eine exogene Variable in die Schätzung mit aufgenommen wird. Diese Multikollinearität könnte durch die Multilevelstruktur

und der damit verbundenen Aggregation begründet sein. Für die Modellierung der Zufallsvariable als kategoriale Variable muss folglich die Stichprobe anders beschaffen sein oder ein anderes Modell zur Anwendung kommen.

Die hier angesprochene Multikollinearität zwischen den unabhängigen Variablen könnte ein Ansatzpunkt für weitere Betrachtungen sein. Nach Albers et al. (2009, S. 224) kann ein Korrelationskoeffizient ab 0,3 bereits auf Multikollinearität hinweisen, wonach eine weitere Untersuchung sinnvoll erscheint. Ein weiterer methodischer Ansatzpunkt für weiterführende Optimierungen ist die Betrachtung der Ausreißer. Im vorliegenden Datensatz konnten 4,4 % der Daten als Ausreißer identifiziert werden, die stichprobenartig auf deren Plausibilität untersucht wurden. Eine umfangreiche Betrachtung oder Eliminierung dieser Daten lässt weitere Verbesserungen des Modells erwarten.

Eine weitere Schwachstelle des Modells ist die mangelnde Signifikanz einiger Variablen, die für die Interpretation wichtig sind. In dieser Arbeit weisen kategoriale Variablen, die als Dummys in das Modell aufgenommen wurden, wie die Präsenz von Haushaltsmitgliedern im Alter von 18 bis 29 Jahren, keine Signifikanz auf. Folglich lässt sich über diese Altersgruppe keine Aussage treffen, weshalb in dieser Arbeit das Ausgangsmodell für die Interpretation herangezogen wurde. Die mangelnde Signifikanz führt dazu, dass keine eindeutige Wirkung ermittelt werden kann und nur eine Tendenz ableitbar ist. Für Informationen über diese Variablen muss eine andere Stichprobe generiert oder ein anderes Modell verwendet werden.

Diese Arbeit zeigt, dass es möglich ist den Pkw-Besitz anhand der Eigenschaften zu beschreiben. Trotz der annehmbaren Güte des Modells gibt es zusätzlichen Optimierungsbedarf, um die Ergebnisse aus dem Modell zu verbessern. Wie bereits erwähnt, lässt die Aufnahme weiterer Variablen, wie beispielsweise für die bauliche Umgebung oder für den sozialen Aspekt, eine Verbesserung des Ergebnisses erwarten. Dabei ist stets der Grundsatz der Sparsamkeit zu beachten. Für aussagekräftigere Ergebnisse wäre es zudem sinnvoll, die Daten auf Haushaltsebene zu erheben, um die Nachteile der Multilevelstruktur und den damit einhergehenden Informationsverlust zu umgehen sowie eine bessere Prognosegüte zu erreichen. Die Nutzung dieser Ergebnisse für die Verkehrs- und Stadtplanung ist nur dahingehend sinnvoll, wenn das Planungsgebiet ähnliche Strukturen wie die in dieser Analyse verwendete Datenstruktur aufweist. Eine Übertragung der Ergebnisse auf ländliche Regionen erscheint nur für ausgewählte Variablen zielführend, da andere strukturelle Voraussetzungen bei den Alternativen gegeben sind. Folglich könnte dieses Modell auf ländliche Regionen ausgeweitet werden. Ähnlich verhält es sich mit anderen Städten im Ausland, da sie sich in gewissen Bereichen durch andere länderspezifische Charakteristika auszeichnen.

Daraus resultieren neue Forschungsfragen, die den Pkw-Besitz in ländlichen Regionen oder anderen Ländern untersuchen können. Es wäre darüber hinaus interessant festzustellen, in welcher Weise sich gewisse Eigenschaften zwischen Stadt und Land bzw. anderen Ländern unterscheiden.

Literaturverzeichnis

Ahrens, G.-A., Ließke, F., Wittwer, R., Hubrich, S. und Wittig, S. (2014). *Methodenbericht zum Forschungsprojekt „Mobilität in Städten – SrV 2013“*. Letzter Zugriff: 20.08.2018. URL: `https://tu-dresden.de/bu/verkehr/ivs/srv/ressourcen/dateien/2013/uebersichtsseite/Methodenbericht_SrV2013.pdf`.

Ahrens, G.-A., Ließke, F., Wittwer, R., Hubrich, S. und Wittig, S. (2015). *Sonderauswertung zum Forschungsprojekt „Mobilität in Städten – SrV 2013“ Stadtgruppe: Große SrV-Vergleichsstädte*. Letzter Zugriff: 20.08.2018. URL: `https://tu-dresden.de/bu/verkehr/ivs/srv/ressourcen/dateien/2013/uebersichtsseite/SrV2013_Stadtgruppe_GrosseSrV-Vergleichsstaedte.pdf`.

Albers, S., Klapper, D., Konradt, U., Walter, A. und J. Wolf (Hrsg.) (2009). *Methodik der empirischen Forschung*. 3. Aufl. Wiesbaden: Gabler.

Backhaus, K., Erichson, B., Plinke, W. und Weiber, R. (2016). *Multivariate Analysemethoden - Eine anwendungsorientierte Einführung*. 14. Aufl. Berlin: Springer Verlag.

Bamberg, G., Baur, F. und Krapp, M. (2012). *Statistik*. 17. Aufl. München: Oldenbourg Verlag.

Bhat, C. R. und Sen, S. (2006). Household vehicle type holdings and usage: an application of the multiple discrete-continuous extreme value (MDCEV) model. In: *Transportation Research Part B: Methodological* 40(1), S. 35–53.

Bundesministerium für Verkehr und digitale Infrastruktur (BMVI) (Hrsg.) (2017). *Verkehr in Zahlen 2017/2018*. Hamburg.

Collin-Lange, V. und Benediktsson, K. (2011). Entering the regime of automobility: car ownership and use by novice drivers in Iceland. In: *Journal of transport geography* 19(4), S. 851–858.

Dienel, H.-L. (2007). Verkehrsgeschichte auf neuen Wegen. In: *Jahrbuch für Wirtschaftsgeschichte* 48(1), S. 19–38.

Maier, G. und Weiss, P. (1990). *Modelle diskreter Entscheidungen - Theorie und Anwendung in den Sozial- und Wirtschaftswissenschaften*. Wien: Springer Verlag.

Schlittgen, R. (2012). *Einführung in die Statistik - Analyse und Modellierung von Daten*. 12. Aufl. München: Oldenbourg Verlag.

Statistische Bundesamt (Destatis) (2018). *Bevölkerung Deutschland 2017*. Letzter Zugriff: 21.08.2018. URL: `https://www.destatis.de/DE/ZahlenFakten/GesellschaftStaat/Bevoelkerung/Bevoelkerungsstand/Tabellen/Zensus_Geschlecht_Staatsangehoerigkeit.html`.

Van Acker, V. und Witlox, F. (2010). Car ownership as a mediating variable in car travel behaviour research using a structural equation modelling approach to identify its dual relationship. In: *Journal of Transport Geography* 18(1), S. 65–74.

Wollschläger, D. (2017). *Grundlagen der Datenanalyse mit R - eine anwendungsorientierte Einführung*. 4. Aufl. Berlin: Springer Spektrum.

Anhang

Übersicht 25 große SrV-Vergleichsstädte

25 Große SrV-Vergleichsstädte	
Augsburg	Kaiserslautern
Berlin	Kassel
Bremen	Kiel
Chemnitz	Leipzig
Cottbus	Magdeburg
Dessau-Roßlau	Mainz
Dresden	Mannheim
Düsseldorf	Potsdam
Erfurt	Rostock
Frankfurt am Main	Schwerin
Gera	Ulm/Neu-Ulm
Halle (Saale)	Zwickau
Jena	

Tabelle A.1: 25 große SrV-Vergleichsstädte (e.D. in Anlehnung an Ahrens et al. (2015, S. 24))

Merkmalsvariablen

Kardinale Merkmalsvariablen
Weglänge
Durchschnittsalter der Volljährigen
Wegeanzahl
Anzahl Personen pro Haushalt
Anzahl motorisierter Zweiräder
Anzahl unmotorisierter Zweiräder
Kürzeste Entfernung zur nächsten ÖV-Haltestelle

Tabelle A.2: Übersicht der kardinalen Merkmalsvariablen (e.D.)

Grafiken der deskriptiven Statistik

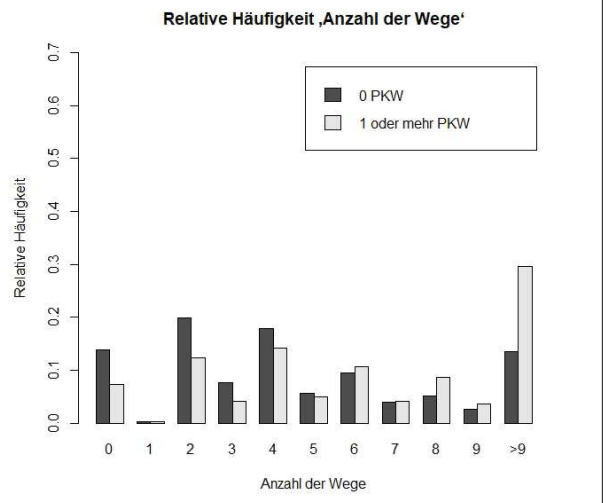

Abbildung A.1: Relative Häufigkeit ‚Anzahl Wege‘ (e.D.)

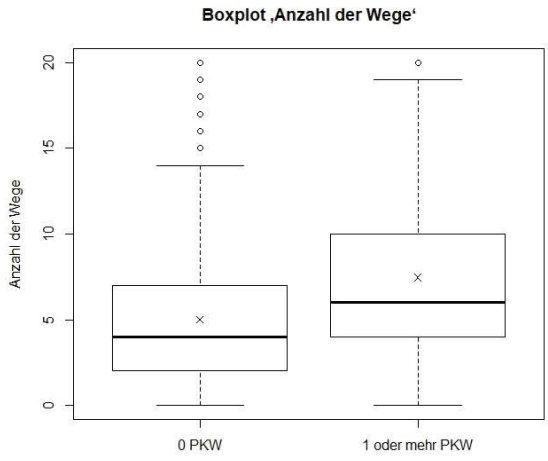

Abbildung A.2: Boxplot der Variable ‚Anzahl Wege‘ (e.D.)

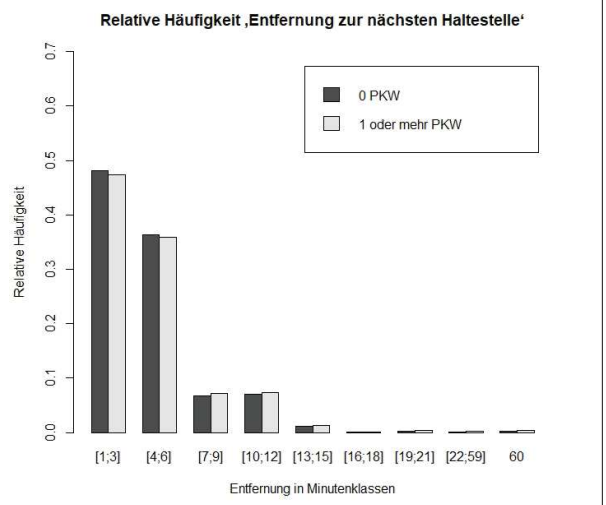

Abbildung A.3: Relative Häufigkeit ‚Weg zur nächsten ÖV-Haltestelle‘ (e.D.)

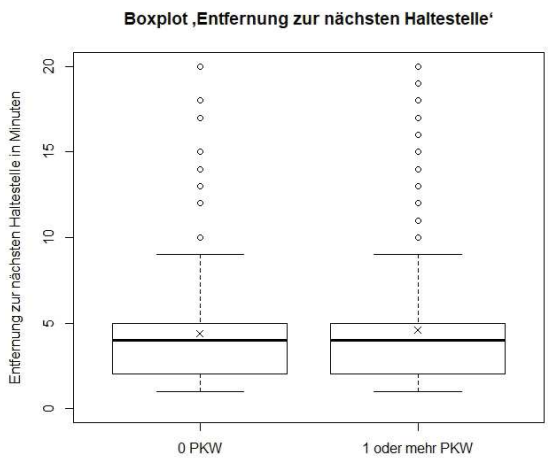

Abbildung A.4: Boxplot der Variable ‚Weg zur nächsten ÖV-Haltestelle‘ (e.D.)

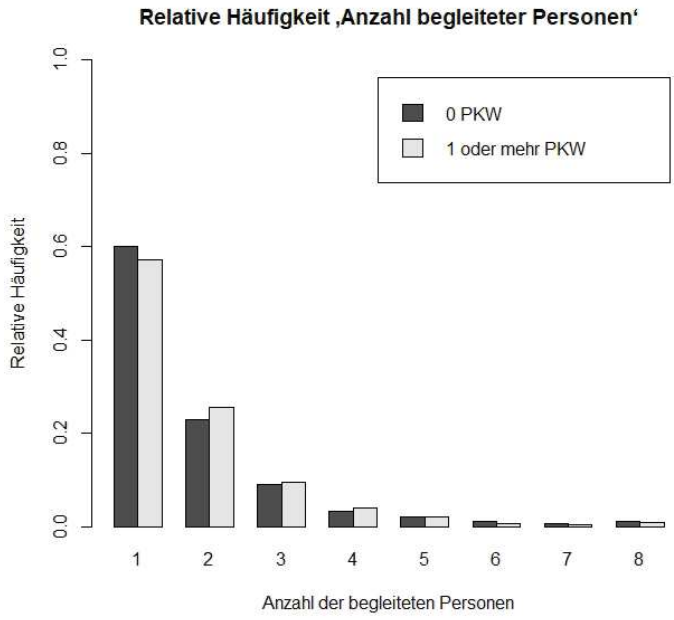

Abbildung A.5: Relative Häufigkeit ‚Anzahl begleitete Personen‘ (e.D.)

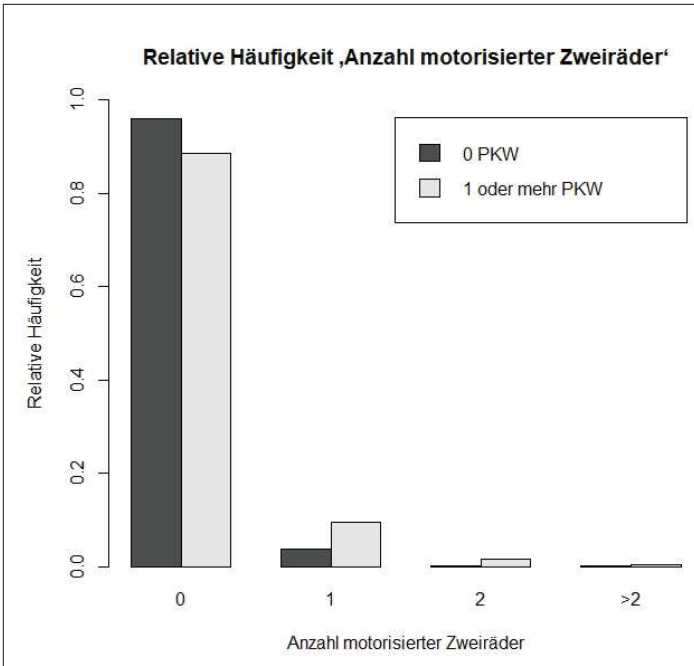

Abbildung A.6: Relative Häufigkeit ‚Anzahl motorisierter Zweiräder‘ (e.D.)

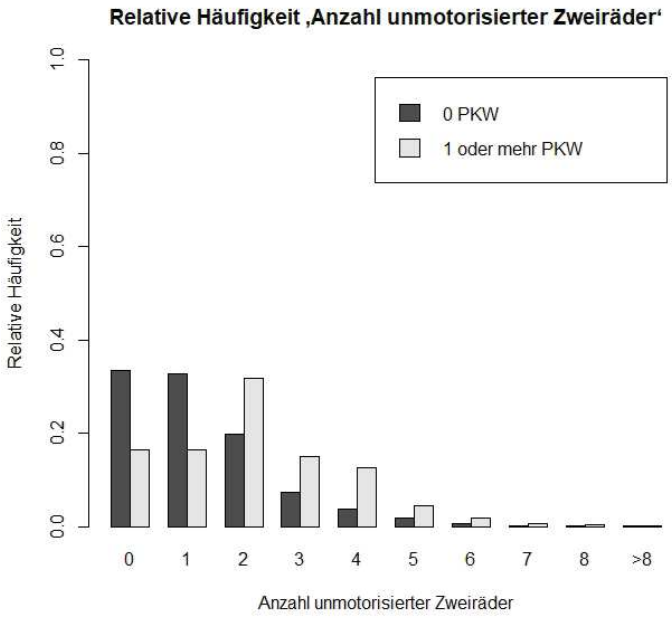

Abbildung A.7: Relative Häufigkeit ‚Anzahl unmotorisierter Zweiräder‘ (e.D.)

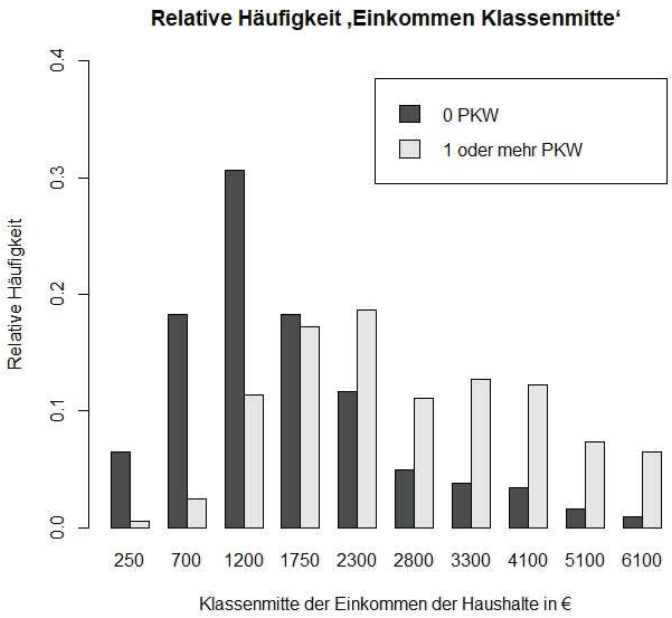

Abbildung A.8: Relative Häufigkeit der Variable ‚Einkommen der Haushalte‘ (e.D.)

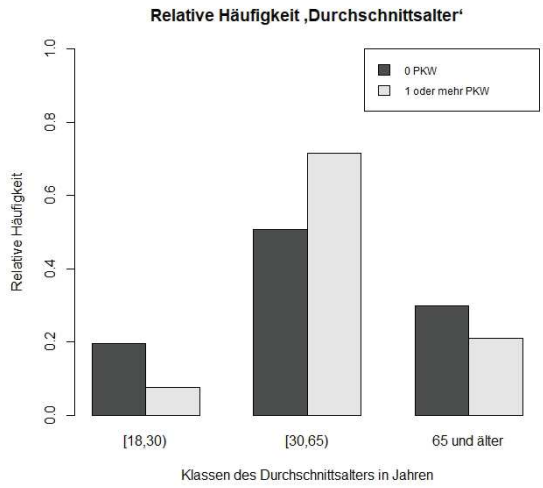

Abbildung A.9: Relative Häufigkeit ‚Durchschnittsalter‘ (e.D.)

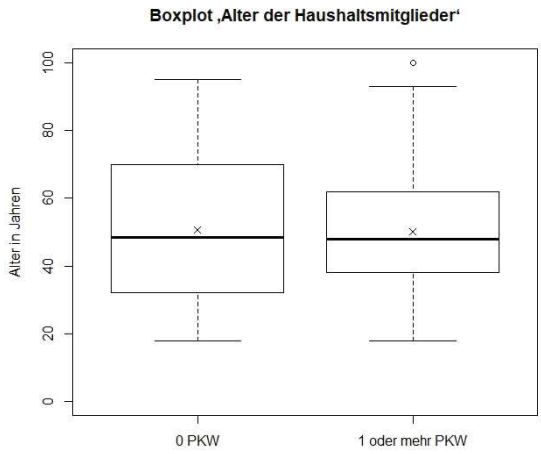

Abbildung A.10: Boxplot der Variable ‚Durchschnittsalter‘ (e.D.)

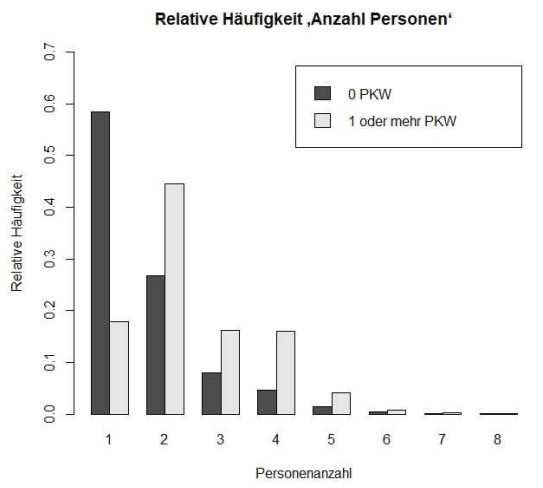

Abbildung A.11: Relative Häufigkeit ‚Anzahl Personen‘ (e.D.)

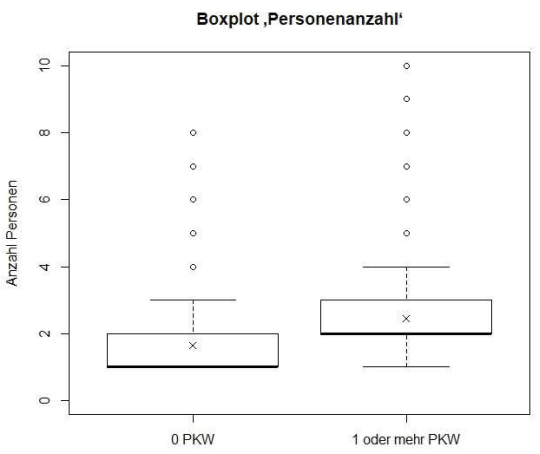

Abbildung A.12: Boxplot der Variable ‚Anzahl Personen‘ (e.D.)

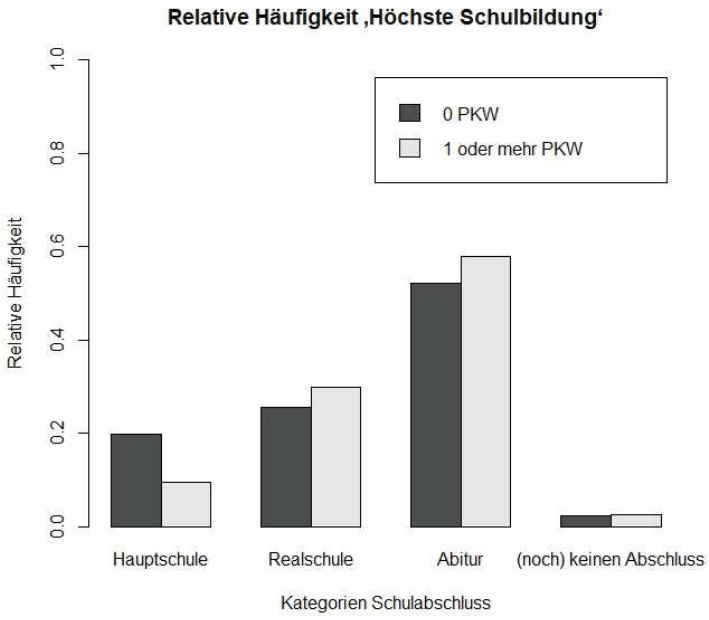

Abbildung A.13: Relative Häufigkeit ‚Höchster Schulabschluss‘ (e.D.)

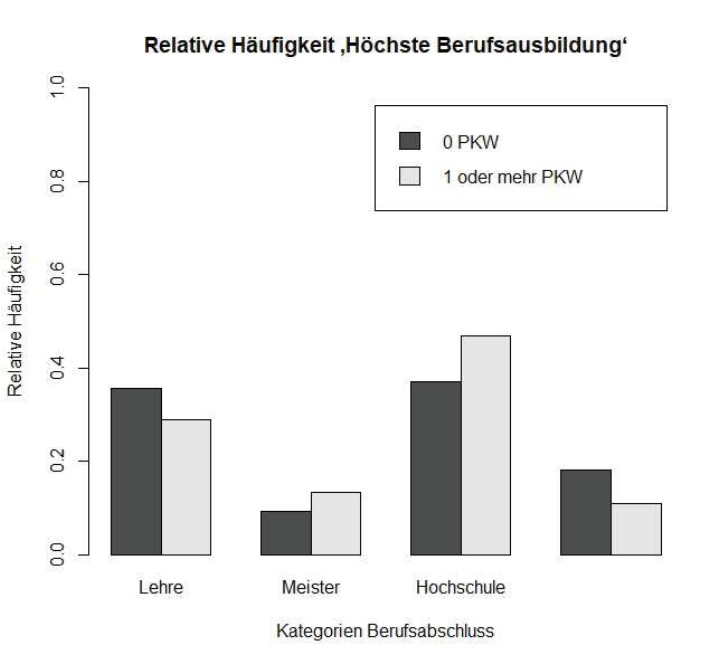

Abbildung A.14: Relative Häufigkeit ‚Höchste Berufsausbildung‘ (e.D.)

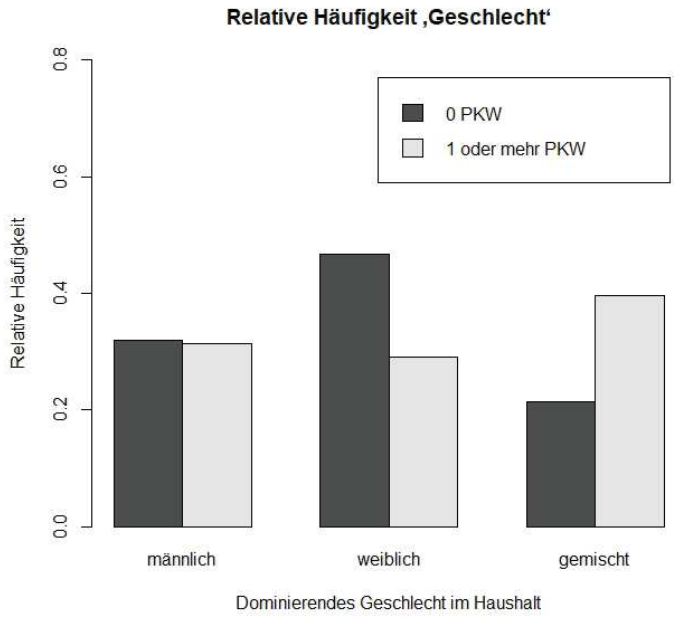

Abbildung A.15: Relative Häufigkeit ‚Geschlecht' (e.D.)

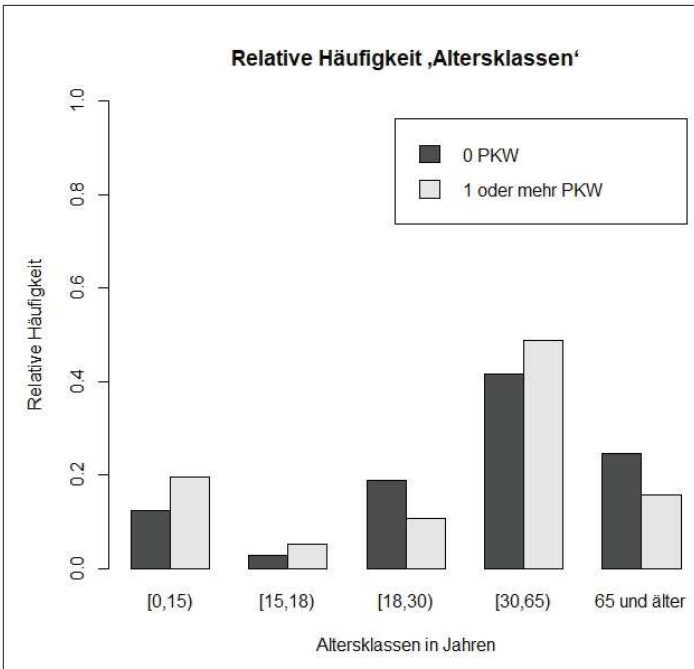

Abbildung A.16: Relative Häufigkeit ‚Altersklassen' (e.D.)

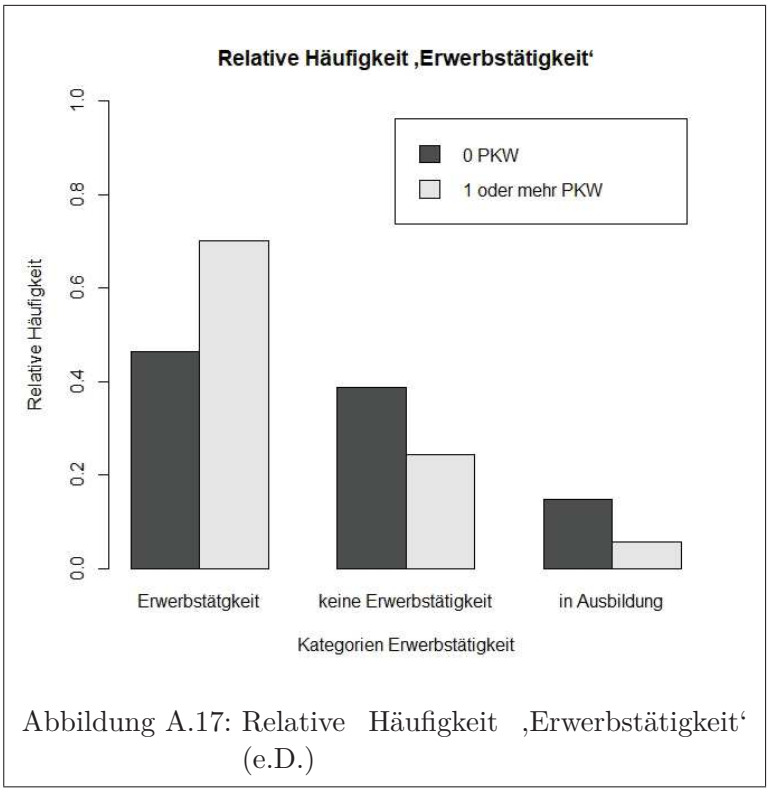

Abbildung A.17: Relative Häufigkeit ‚Erwerbstätigkeit' (e.D.)

Korrelationstabelle für metrische Variablen

	Anzahl Personen	Anzahl motorisierter Zweiräder	Anzahl unmotorisierter Zweiräder	Entfernung zur nächsten Haltestelle	Einkommen	Durchschnittsalter	Wegeanzahl	Anzahl begleiteter Personen	Wegelänge
Anzahl Personen	1								
Anzahl motorisierter Zweiräder	0,119	1							
Anzahl unmotorisierter Zweiräder	0,637	0,170	1						
Entfernung nächste Haltestelle	-0,039	-0,011	-0,065	1					
Einkommen	0,429	0,109	0,414	-0,051	1				
Durchschnittsalter	-0,313	-0,102	-0,253	0,131	-0,144	1			
Wegeanzahl	0,562	0,070	0,440	-0,059	0,259	-0,229	1		
Anzahl begleiteter Personen	0,272	0,028	0,198	-0,019	0,114	-0,127	0,333	1	
Weglänge	0,267	0,087	0,216	0,002	0,210	-0,172	0,378	0,161	1

Tabelle A.3: Korrelationstabelle metrischer Merkmalsvariablen (e.D.)

Korrigierter Kontingenzkoeffizient für ordinal skalierte Variablen

Kategoriale Variable	Ausprägungen	K_*
Höchster Schulabschluss	(Noch) ohne Abschluss (Referenzkategorie) Hauptschulabschluss Realschulabschluss Abitur	0,183
Höchste Berufsausbildung	(Noch) ohne Abschluss (Referenzkategorie) Lehre Meister Hochschulabschluss	0,171
Geschlecht	gemischt (Referenzkategorie) weiblich männlich	0,242
Erwerbstätigkeit	in Ausbildung (Referenzkategorie) erwerbstätig nicht erwerbstätig	0,043
Altersklassen	0 bis 14 Jahre 15 bis 17 Jahre 18 bis 29 Jahre 30 bis 65 Jahre über 65 Jahre (Referenzkategorie)	0,245

Tabelle A.4: Korrigierter Kontingenzkoeffizient für kategoriale Variablen (e.D.)

Variablen für das binäre Logit Ausgangsmodell

Variable	**Skalierung**
Anzahl Personen	kardinal
Anzahl motorisierter Zweiräder	kardinal
Kürzeste Entfernung zur nächsten ÖV-Haltestelle	kardinal
Einkommensklassen	pseudokardinal
Durchschnittsalter der Volljährigen	kardinal
Wegeanzahl	kardinal
Wegelänge	kardinal
Verfügbarkeit Dienstwagen	nominal
Einschränkung	nominal
Pkw-Führerschein	nominal
Sonstiger Führerschein	nominal
Besitz Dauerkarte	nominal
Nutzung Carsharing	nominal
Begleitung einer Person	nominal
Altersklassen	ordinal
Höchster Schulabschluss	kardinal
Höchste Berufsausbildung	kardinal
Geschlecht	ordinal

Tabelle A.5: Ausgangsvariablen für das binäre Logit-Modell (e.D.)

Übersicht Ergebnisse der Güteprüfung

Gütemaß	**Wert für das endgültige Logit-Modell**
AIC	11.340
BIC	11.518
LL-Wert	-5.647
LLR	6.154,4 | p-Wert 0,000
Mc-F R^2	0,353
R^2_{CS}	0,307
R^2_N	0,475
Trefferquote	0,789 > 0,658 (PCC)
AUC	0,873

Tabelle A.6: Übersicht der Ergebnisse der Güteprüfung für das endgültige Modell (e.D.)

Vergleich kategorialer Variablen aus dem Ausgangsmodell mit einem stark reduzierten Modell

		Odds-Ratio	Odds-Ratio	Relatives Risiko	95 % Konfidenzintervall Ausgangsmodell	
		Ausgangs-modell	kleines Modell	Ausgangs-modell	von	bis
Altersklassen	0-14 Jahre	0,458	0,424	0,973	0,372	0,563
	15-17 Jahre	0,528	0,498	0,976	0,397	0,710
	18-29 Jahre	1,001	0,954	1,000	0,779	1,289
	30-65 Jahre	1,076	1,049	1,002	0,920	1,258
	über 65 Jahre	Referenzkategorie				
Schulausbildung	Hauptschulabschluss	0,752	0,762	0,991	0,473	1,197
	Realschulabschluss	1,010	1,042	1,000	0,638	1,575
	Abitur	0,617	0,646	0,987	0,394	0,953
	(Noch) keinen Abschluss	Referenzkategorie				
Berufsausbildung	Lehre	1,424	1,461	1,010	1,264	2,101
	Meisterabschluss	1,628	1,658	1,012	1,155	1,755
	Hochschulabschluss	1,074	1,093	1,002	0,871	1,325
	(Noch) keinen Abschluss	Referenzkategorie				

Tabelle A.7: Ergebnisse kategorialer Variablen (e.D.)

Konfidenzintervalle

	95 % Konfidenzintervall Odds-Ratio	
	von	bis
Interzept	0,016	0,034
Anzahl der Personen	2,100	2,428
Verfügbarkeit Dienstwagen	» 1	» 1
Entfernung nächste Haltestelle	1,005	1,026
Einkommen	1,001	1,001
Durchschnittsalter	1,002	1,010
Anzahl motorisierter Zweiräder	1,273	1,838
Einschränkung	0,505	0,697
Hauptschulabschluss	0,609	0,887
Abitur	0,527	0,725
Meisterabschluss	1,270	1,909
Lehre	1,174	1,601
Pkw-Führerschein	13,008	19,204
Sonstiger Führerschein	1,376	1,841
Besitz Dauerkarte	0,247	0,323
Nutzung Carsharing	0,121	0,184
Wegeanzahl	0,965	0,995
Altersklasse 0 bis 14 Jahre	0,386	0,569
Altersklasse 15 bis 17 Jahre	0,403	0,718
Geschlecht männlich	0,725	0,979
Geschlecht weiblich	0,645	0,864
Begleitung von Personen	1,056	1,429
Wegelänge	1,010	1,014

Tabelle A.8: Konfidenzintervalle des endgültigen Modells (e.D.)

Danksagung

Ich danke der Hanns-Seidel-Stiftung, die mich als Stipendiat vom Oktober 2016 bis Januar 2018 finanziell mit Mitteln des Bundesministeriums für Bildung und Forschung und bis zum Ende meines Bachelorstudiums ideell unterstützt hat. Zudem bedanke ich mich bei der Studienstiftung des deutschen Volkes, die mich als Stipendiat seit Februar 2018 ideell und finanziell unterstützen.

Selbstständigkeitserklärung

Ich erkläre hiermit, dass die vorliegende Arbeit selbstständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt wurde. Die aus fremden Quellen wörtlich oder sinngemäß übernommenen Gedanken sind als solche kenntlich gemacht. Ich erkläre ferner, dass ich die vorliegende Arbeit an keiner anderen Stelle als Prüfungsarbeit eingereicht habe oder einreichen werde.

Dresden, 06. September 2018

_____

Stefan Martin Lins